DOCUMENT RESUME

ED 193 282 TM 800 597

AUTHOR Elliett, Frederick S., Jr.
TITLE Fairness and the Predictors.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Apr 80
NOTE 101p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *College Admission; *College Entrance Examinations; *Culture Fair Tests; Higher Education; *Predictive Validity; *Selection; *Test Bias; Testing Problems; Test Validity

ABSTRACT

Certain conditions are investigated as to whether they are necessary or sufficient to indicate the fairness of a test, or predictor, in its role in the college admissions procedure. Eight general positions are considered that have claimed to provide sufficient conditions for the unbiasedness of a standardized test or predictor: (1) equal test score means for the relevant groups; (2) equality of prediction systems; (3) equal risk (Einhorn and Bass); (4) constant ratio of acceptance for different groups (Thorndike); (5) equal conditional probability of selection (Cole); (6) equal probability of selection; (7) a culture-modified criterion (Darlington); and (8) equal expected utility (Petersen and Novick). These positions are described, and it is argued that none of the positions provides a sufficient condition for fairness; that only two provide conditions which are necessary for fairness in special situations; and that current selection policies using standardized tests are only fair in the grossest sense. A policy is recommended that uses the "best available predictors," but in a way which is reasonably efficient and more fair than current practices. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED193282

FAIRNESS AND THE PREDICTORS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
F. ELLett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

Frederick S. Ellett, Jr.

University of California, Los Angeles

March 20, 1980

TM 800597

Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980

The research reported herein was supported in whole or in part by a grant to the Center for the Study of Evaluation from the National Institute of Education, U. S. Department of Health, Education, and Welfare. However, the opinions and findings expressed here do not necessarily reflect the position or policy of NIE and no official NIE endorsement should be inferred.

Fairness and the Predictors

FREDERICK S. ELLETT, Jr.

FREDERICK S. ELLETT, Jr.
405 Hilgard Avenue
University of California
Los Angeles, CA 90024

Summary

In the last decade or so several educational researchers and psychometricians have devoted themselves to the question whether the use of standardized tests or predictors in college admissions procedures is fair to certain ethnic and racial minorities. They have talked about this question in terms of test bias and fairness (Anastasi, 1968; Bass, 1976; Bass & Einhorn, 1971; Breland, 1973; Breland & Ironson, 1976; Cleary, 1968; Cole, 1973; Cronbach, 1976; Darlington, 1971, 1976; Guion, 1966; Hunter & Schmidt, 1976; Linn, 1973, 1975a, 1975b, 1976; Linn & Werts, 1971; Novick & Ellis, 1977; Novick & Petersen, 1976; Petersen & Novick, 1974, 1976; Sawyer, Cole, & Cole, 1976; Schmidt & Hunter, 1974; and Thorndike, 1971).

From these writings one can find eight general positions which claim to provide a sufficient condition for the fairness of a test or a predictor in its role in a selection procedure. These are: (1) the equal test score means for the relevant groups condition, (2) the equality of prediction systems (or regression) condition, (3) the Einhorn and Bass equal risk condition, (4) the Thorndike equal constant ratio condition, (5) the Cole equal conditional probability condition, (6) the equal probability condition, (7) the Darlington culture-modified criterion condition, and (8) the Novick and Petersen decision-theoretic (or equal expected utility) condition. In this essay, these positions are described, and it is argued that none of the positions provides a sufficient condition for fairness and that two of the positions provide a necessary condition for fairness only in very special situations.

It is concluded that current procedures which use standardized tests are fair in only the grossest sense. Further, a policy is recommended which is reasonably efficient but more fair than current procedures.

It is assumed that the institutions have good reasons for pursuing their (global) purposes (Assumption I). Also, it is assumed that the institutions have objectively determined the qualifications which students must have in order for these institutions to achieve their global purposes, and that these (theoretical) qualifications have a reliable, unbiased, perfectly relevant criterion measure which can be operationally defined as the first-year grade point averages (Assumption II). Although these assumptions need justification, it is shown that the selection procedure is liable to criticism at another aspect; that aspect which is concerned with the role of the empirical predictors of criterion performance.

The paper draws the distinction between procedural fairness and substantive fairness. It suggests that there are three substantive principles of fairness which are applicable to college selection procedures: the principle of equal chance for all, the principle of equal opportunity for those suitably qualified, and the derivative principle of equal accuracy.

It is argued that none of the positions states a sufficient condition, that the Thorndike condition is a necessary condition according to the principle of equal chance for those suitably qualified in special situations, and that the Cole condition is a necessary condition according to the principle of equal accuracy in special situations. Given Assumptions I and II, then it is argued that educational researchers and psychometricians have not captured the essential conditions for a test to be fair in its role in the selection procedure.

It is also argued that current selection policies which are based upon the "best available" prediction equations are remarkably efficient but that they are fair in only the grossest sense according to all three of the substantive principles of fairness. Whether or not following current practices can be shown to be morally right, it is important to recognize that these policies are unfair to both minority and majority applicants. Further, a policy is recommended which uses the "best available" predictors, but which is reasonably efficient and more fair than the current policies.

One final point. This paper has been investigating whether certain conditions are necessary or sufficient for the fairness of the test (or empirical predictor) in its role in the selection procedure. It has been investigating whether a particular kind of unfairness may correctly be attributed to the test in its role in the selection procedure. One must remember that there are other kinds of unfairness that a selection procedure can bring about. It is most important that we determine whether the institution's purposes and its related qualifications are themselves just. And here we must be concerned with questions of compensatory justice and with questions about the institution's role in the distribution of social and economic benefits and burdens. Such questions belong to the realm of political and moral philosophy and, at the present, the only method we have for dealing with such questions is moral deliberation and debate. There is no moral calculus, no moral geometry; moral debate relies on moral reasoning and judgment.

5:FSE

Fairness and the Predictors

FREDERICK S. ELLETT, Jr., U.C.L.A.

Abstract

Eight general positions are considered that have claimed to provide a sufficient condition for the fairness or unbiasedness of a (standardized) test or predictor in its role in the selection procedure. It is argued (a) that none of these positions provides a sufficient condition for fairness; (b) that only two provide conditions which are necessary for fairness in very special situations; and (c) that current selection policies using standardized tests are only grossly fair. A policy is recommended that uses the "best available predictors," but in a way which is reasonably efficient and yet more fair than the current practices.

Abstract

This article considers eight general positions that have claimed to provide a sufficient condition for the fairness of a (standardized) test or predictor in its role in the selection procedure. It is argued that none of these positions provides a sufficient condition for fairness; and that only two provide conditions which are necessary for fairness in very special situations. It is argued that current selection policies that use standardized tests are fair in only the grossest sense. A policy is recommended that uses the "best available predictors," but that uses the test in a way that is reasonably efficient and yet more fair than the current practices. (It is pointed out how the justification of such a policy in general depends on further ethical principles.)

Introduction

In the last decade or so several educational researchers and psychometricians have devoted themselves to the question whether the use of standardized tests in college admissions procedures is fair to certain ethnic and racial minorities. They have talked about this question in terms of test bias and fairness (Anastasi, 1968; Bass, 1976; Bass & Einhorn, 1971; Breland, 1973; Breland & Ironson, 1976; Cleary, 1968; Cole, 1973; Cronbach, 1976; Darlington, 1971, 1976; Guion, 1966; Hunter & Schmidt, 1976; Linn, 1973, 1975a, 1975b, 1976; Linn & Werts, 1971; Novick & Ellis, 1977; Novick & Petersen, 1976; Petersen & Novick, 1974, 1976; Sawyer, Cole, & Cole, 1976; Schmidt & Hunter, 1974; and Thorndike, 1971).

From these writings one can find eight general positions which claim to provide a sufficient condition for the fairness of a test in its role in a selection procedure. These are: (1) the equal test score means for the relevant groups condition, (2) the equality of prediction systems (or regression) condition, (3) the equal risk condition, (4) the Thorndike equal constant ratio condition, (5) the Cole equal conditional probability condition, (6) the equal probability condition, (7) the culture-modified criterion condition, and (8) the decision-theoretic (or equal expected utility) condition. In this essay, I will describe the positions. I will argue that none of the positions provides a sufficient condition for fairness, and I will argue that two of the positions provide a necessary condition for fairness only in very special situations.

I will conclude that current procedures which use standardized tests are fair in only the grossest sense. Finally, I recommend a policy which is reasonably efficient but more fair than current procedures.

The work by educational researchers and psychometricians has occurred during a period in which many institutions of higher education have been concerned about increasing the number of students from certain racial and ethnic minorities. Some writers have presented arguments which purport to show that increasing the number of students from these groups is fair because compensatory justice (or social justice) requires that the traditional set of qualifications for admissions be overridden. Some writers have presented arguments which purport to show that increasing the number of students from these groups is fair because the institutions really are (or should be) committed to achieving purposes whose set of relevant qualifications, when administered properly, would result in this increase. The literature dealing with the moral debate of the justice of preferential treatment (or reverse discrimination) is growing rapidly (Blackstone, 1975; Boxill, 1978; Fiss, 1976; Fullinwider, 1975; Heslep, 1976; Goldman, 1975, 1976, 1977; Gross, 1977; Jagger, 1977; Nagel, 1973; O'Neil, 1975; Sher, 1975; Simon, 1974; Strike, 1976; Thomson, 1973; Wasserstrom, 1976, 1977). Two of perhaps the most controversial legal cases--the Defunis case and the Bakke case--were concerned with the justice of the policies of institutions of higher education.

Depending on the ethical validity of these arguments, a certain set of educational purposes and their related qualifications would be partially delimited. As I intend to show, the question of whether certain conditions are necessary or sufficient for the fairness of a test in its

role in the selection procedure is logically independent of whether the educational purpose and related qualifications are justified.

Whether unfairness is brought about by the empirical indicators at one aspect of the selection process is logically independent of whether unfairness is brought about at other aspects of the selection procedure. It could happen that compensatory justice is properly attended to, but that unfairness is brought about or sustained by the empirical indicators; also, it could happen that the empirical indicators cause no unfairness, but that the administration has been unfair by disregarding justifiable claims based on compensatory or social justice, or by failing to pursue fully the institution's legitimate purposes. As I intend to show there is a strong sense in which the unfairness of the selection procedure can be traced to characteristics of the test.

It is extremely difficult to get clear about the issues involved in test fairness for the discussion must go on at a time when various minority and ethnic groups are claiming that society is treating them unjustly or unfairly with respect to distribution of economic and social benefits. Some of these groups have charged that institutions of higher education currently employ predictors (or indicator standards) that are unfair to members of certain racial or ethnic minority groups. It has been claimed that these predictors do not predict minority applicants as accurately as majority members and that they underpredict the performance of minority applicants. Partially in response to these claims, educational researchers and psychometricians have become interested in the question as to whether current predictors are unfair to certain racial or ethnic minorities. In these issues one has to distinguish between the charge that institutions of higher education are pursuing

purposes that result in economic and social injustice and the charge that tests play a role in the selection procedure which leads to unfairness of another kind.

In this essay, then, I will explicate and criticize eight general kinds of conditions which the educational researchers have put forward. I will argue that none of these provides a sufficient condition for fairness, and that two of these are necessary conditions for fairness only in special situations. I will argue that current selection procedures that use standardized tests are fair in only the grossest sense. I will recommend a policy that uses the "best available predictors," but that is reasonably efficient and yet more fair than the current practices.

Kinds of Fairness: Assumption I

It is important to recognize that there are three kinds of unfairness which can occur in the selection process at institutions of higher education (Feinberg, 1973, pp. 103-107). In the first kind, the administrator may act irregularly--with partiality, with prejudice, or bias--so that he does not base his decision solely on whether an applicant satisfies the necessary qualifications for admissions. In other words, in the first kind of fairness, the administrator acts fairly if, and only if, he properly follows the procedure or guidelines. When an administrator acts in this way, he is said to be procedurally fair.

Procedural fairness, however, has only conditional value. Whether it has moral value depends upon the fairness of the results which are brought about by following the procedure or guidelines. And in the second kind of unfairness, the rules and guidelines by which an admissions officer selects applicants may discriminate against kinds of people who

differ in respects that are not relevant to the agreed-upon objectives that lie behind student selection. For example, in formulating or operationalizing these qualifications, the institution may make mistakes, or fall short of objectivity and impartiality, with the result that the formulation may be unfair--granting undeserved advantages in selection to certain kinds of people and arbitrarily imposing disadvantages upon other kinds of people.

Thus, the moral value of procedural fairness is logically dependent upon the fairness of the results which it brings about and the fairness of the results is logically dependent upon the purposes of the institution. There is also a third kind of unfairness which can occur. The purpose of the educational institution may be arbitrary or invidious--in short, unfair. For example, the institutions may be pursuing (intentionally or unintentionally) a purpose which is unjust, such as the conservation of economic privilege for a favored class, the "legitimating" of an arbitrary rationing of privilege, or the maintaining of racial segregation within society. If the purpose which the institution is pursuing is unjust, then the moral value of doing things properly in the first way and in the second way is undermined. One is not entitled to much moral praise if he impartially administers a rule which itself is unjust; indeed, he may be subject to severe moral censure.

The purposes which an institution pursues must be, at least, not unfair and not immoral. Furthermore, since the institutions of higher education do have important effects on the lives of their applicants, among others, the institutions should have good reasons for pursuing certain purposes. Now one should not take lightly the charges that institutions of higher education are pursuing unjust purposes. Indeed,

these charges should be given the most careful attention. Although more criticism could profitably be spent on these matters, I shall leave them aside. In this essay, then, I shall assume that good reasons related to fairness can be provided for the institution's pursuing its purposes (Assumption I). I want to show that the institution is still liable to serious criticism at another aspect of the selection process when it selects students by means of standardized tests.

Theoretically Relevant Qualifications: Assumption II

Also, I shall assume that the institution has objectively determined what the qualifications of its students should be in order to achieve its purposes. It is convenient to call these qualifications the theoretically relevant qualifications (Heslep, 1976, p. 147).

Part of the job of the educational researcher is to find or develop practical, empirical measures of the theoretical qualifications. These empirical measures are often referred to as criterion measures. Such measures, of course, may fall short of the ideal. The criterion measure may be an unreliable indicator of the theoretical qualifications, it may be a biased indicator, or it may be influenced by factors unrelated to the theoretical qualifications. Most of the studies on the fairness of educational selection processes have taken the criterion measure to be the student's first-year grade point average. And it is not wildly implausible to assume that colleges and universities would effectively carry out their purposes if they had successful students. At any rate, they have implicitly assumed that first-year grade point average is a fairly reliable, unbiased, and accurate measure of the theoretical qualifications. Clearly, without some convenient and objective way of measuring the theoretical qualifications, little large-scale, objective

evaluation can be done on whether the selection process is efficient or fair.

Now, more needs to be said about the adequacy of the first-year grade point average (FYGPA) as a criterion measure, and this, in turn, must rely on an adequate delimitation of theoretical qualifications. I believe that both of these matters should be given serious attention, but I shall leave them aside. In this essay, I shall assume that the theoretical qualifications have been objectively determined and defended and that the criterion measure (i.e., FYGPA) is a reliable, unbiased, perfectly relevant indicator of the theoretical qualifications (Assumption II). Educational researchers are not unaware of these difficulties, but they make similar assumptions in defining test fairness (Thorndike, 1971, pp. 69-70; Linn, 1975b, p. 4). I intend to show that using standardized tests still leads to a serious kind of unfairness.

In the ideal case, the admissions officer would predict exactly an applicant's FYGPA and choose the applicants who are predicted, say, to have the highest FYGPA. An admissions officer would reach his decisions on the basis of the characteristics which the applicants have and on the basis of the relationship between those characteristics and FYGPA. Perhaps those characteristics would include the person's grade point average at previous institutions; perhaps they would include the person's score on standardized tests which were empirically designed to predict FYGPA. The characteristics which are empirically related to the criterion measure FYGPA are called indicator <u>standards</u> or <u>empirical predictors</u>. Questions of test fairness are concerned with the role of the empirical predictors in the selection procedure.

If such ideal empirical predictors were to exist, then the possibilities of unfairness of the first and second kind would be practically eliminated. With respect to the first kind, the administrator would have a procedure or set of guidelines to follow in making his decision, and other people could check on him to make sure his acts are procedurally fair. With respect to the second kind, since one could make exact predictions, one would be selecting students on the basis of the criterion measure which, in turn, is perfectly related to the theoretical qualifications. In this ideal situation, students would be selected on the basis of the relevant qualifications. Thus, the possibility of unfairness at the second level would be eliminated. Given the assumption about the legitimate purposes of the institution, I shall refer to fairness of the second kind as substantive fairness.

Principles of Substantive Fairness

In the case of college admissions, the first-year grade point average is taken as the criterion measure of the theoretical qualifications and the relevant qualifications can be expressed in terms of first-year grade point averages. What is substantially fair in one situation, of course, may be different from what is fair in another situation, for what is fair depends in part on what purposes the institution is pursuing. Consider Situation 1. Person A should be preferred over Person B if, and only if, A's first-year grade point average is greater than that of B. Therefore, those with the highest qualifications are chosen first; an applicant is only admitted after those with higher qualifications have been admitted. Fairness requires that everyone with the same qualifications should be treated the same. Each and every person at a given qualification level should be treated in the same

way--either rejected or accepted. Furthermore, those applicants with the higher qualifications should be accepted before those with the lower qualifications.

When there are more applicants than openings, fairness would require, as before, that those with the highest qualifications be admitted first and that equals be treated equally. There may come a time, however, when the number of those at a given qualification level is greater than the number of openings which remain. At this point, fairness would require that the openings be distributed to those applicants by lot. I shall call the principle of fairness, which is characteristic of Situation 1 under scarcity, the principle of proportional equality.

Situation 2 is the following. The range of the criterion measure is divided into a finite number of sets. In each of the sets, the people whose scores fall into the set are regarded as equals. Furthermore, there is a rank ordering among the sets so that the highest is preferred over the next highest set, and so on. For example, it may be that the college, after considering its purposes, can defend the construction of two relevant sets (or groups) which are defined in the following way: (1) all of those with a first-year grade point average above Y belong to the upper group and (2) all of those with averages less than or equal to Y belong to the lower group. Depending on the value of Y, the high group may contain applicants who are highly qualified, or it may contain applicants who are at least minimally qualified (i.e., they could do the work). (Of course, the institution will decide to use a grouping which best suits its purposes. It may, for example, be able to defend using a five-group system.) Fairness requires that everyone with the same qualifications should be treated the same.

Therefore, if there are enough openings, all of those in the upper group should be admitted before those in the lower group. If there are more applicants in the upper group than there are openings, then fairness would require that the openings be distributed by lot. I shall call the principle of fairness, which is characteristic of Situation 2 under scarcity, the principle of equal opportunity for those suitably qualified.

In certain situations, it is plausible to argue that there is a derivative principle of fairness. Suppose that a finite number of groups can be shown to be related to the institution's purposes, as in Situation 2. In Situation 3, however, it is recognized that errors will be made in the selection procedure and, therefore, that applicants will be misclassified and hence given the wrong treatment. Although errors will be made, it is possible to distribute the errors in such a way that each person will be predicted to the same degree of accuracy. In other words, if a person were suitably qualified, then he would be predicted with probability P to be suitably qualified. If a person were not suitably qualified (supposing a two-group system), then he would be predicted with probability P to be not suitably qualified. Although mistakes in prediction will be made, every person has the same risk of being misclassified. The risk is distributed equally to all individuals. Therefore, it is plausibly claimed that fairness would require that the error be equally distributed. I shall call this derivative principle of fairness, which is characteristic of Situation 3, the principle of equal accuracy (I shall have more to say about it later on). In a sense, this principle supplements the previous principles of fairness.

Here it is appropriate to point out the nature of the criticism that I advance. Educational researchers and psychometricians have

offered various definitions of the fairness of the role of the tests in the selection procedure. I intend to show that under two general assumptions (I, II) the definitions provide neither necessary nor sufficient conditions for fairness. The criticism is especially forceful because it relies on the very assumptions usually made by psychometricians and because such assumptions would seem to be defensible in certain, actual cases. Furthermore, whatever the plausibility of the assumptions for actual cases, my argument which is built upon these assumptions and which shows that unfairness results, shows that the psychometricians' definitions have not captured the essential conditions for the fairness of the empirical predictors in their role in the selection procedure.

Optimistic Views About Predictions

At times, educational researchers have been quite optimistic about the role which predictors (and especially tests) might play. Here is a typical remark made by A. Anastasi (1968):

> When social stereotypes and prejudice may distort interpersonal relations, tests provide a safeguard against favoritism and arbitrary or capricious decisions. (p. 563)

If this statement says only that using a test assures procedural fairness, then the statement is true but of little importance. Procedural fairness is of value insofar as it is conducive to substantive fairness. On the other hand, Anastasi could be understood as saying that tests provide a safeguard against unfairness at the first and second levels. If tests were to assure this, then their use would be of great value.

In summarizing her discussion on the objectivity of tests, Anastasi wrote:

> Tests can be misused in testing culturally disadvantaged persons, as in testing anyone else. When properly used, however, they serve an important function in preventing irrelevant and unfair discrimination (p. 564)

In such remarks, it seems that Anastasi maintained that the use of predictors (or tests) would safeguard fairness of the first and second kinds. If tests were not to safeguard substantive fairness, but were to cause unfairness of the second kind, then there would be little or no moral value in being procedurally fair (i.e., securing fairness of the first kind), for the value of procedural fairness is logically dependent upon the fairness of the results which are brought about by properly following the procedure. If there were unfairness of the second kind, then an administrator could not justify his behavior merely by showing that he has achieved procedural fairness, for he is in effect discriminating against people on irrelevant grounds.

The Nature of Prediction: The Task

The admissions officer would like to have a practical procedure by which he would predict an applicant's level of success in college, given certain information about the applicant--such as his previous school grades and his scores on certain standardized tests. By collecting data on many applicants over a period of time, the educational researcher may be able to develop a good, linear equation for predicting the criterion. (It has been noted that recent studies of nonlinear models for predicting grades in law school suggest that little, if anything, would be gained by using a nonlinear prediction system (Linn, 1975b, p. 4)). After finding a statistical relationship between the criterion and the predictor, he can find the best linear estimate of the criterion based on the predictor. The linear predictor is best in the sense that the mean of the squares of differences between the estimated value and the actual criterion value (i.e., the error of estimate) is minimized. The linear predictor (or estimate) is a linear function of the predictor.

It turns out that the Pearson product-moment correlation between the criterion and the predictor has special significance in forming the best linear estimates. In particular, if one expresses the results in terms of standard scores, then the best linear estimate of the criterion score (in standard form) is equal to the product of the correlation coefficient and the predictor score (in standard form). This equation is called the regression equation, and this kind of estimate is also called the linear regression estimate. It is the most common form of estimate (or prediction) used today. (When the raw scores are used, it turns out that the slope of the regression line is proportional to the value of the correlation coefficient.)

Empirical Relationships: Correlations

The degree to which the predictor and performance are corrrelated, then, is expressed in terms of a number called the correlation coefficient. The possible range of values that the correlation coefficient may take is -1.00 to +1.00 although, for predictions of success in school, all commonly used predictors correlate positively with outcomes. In the ideal case, depicted above, where the administrator can predict precisely the applicant's first-year grade point average, the value of the correlation coefficient is 1.00. In situations where certain levels of performance are being predicted, the correlation coefficient is also called the predictive validity.

Now, there are some general important facts about predictors which should be understood. Empirical research has shown that, generally, whatever prediction of college accomplishment which either previous school grades or test scores give, the two in combination will give better predictions (American College Testing Assessment Program, 1973,

p. 150). Also, while generally good prediction is possible, there is variability among colleges on the degree of prediction attained. For example, in a 1968 sample, the correlations between certain standardized tests and the first-semester grade point average varied from .29 to .84 with a median value of .45; in a 1969 sample, the values varied from .33 to .60 with a median value of .47 (p. 138). In the interests of fairness and efficiency, then, an institution should collect and study the empirical data concerned with its own selection procedure. It has also been shown that there is a decline in the correlation between certain tests and the eight college semester grades from .47 to .21 (p. 145). Using standardized tests (e.g., the American College Testing Program (ACT) tests) and the high school GPA as predictors raises the correlation with criterion performance (i.e., FYGPA) to a median value of about .58 (p. 158). Graduate school performance prediction is usually less efficient. In law school selection, the weighted combination of the Law School Admission Test (LSAT) and undergraduate grade point average and, perhaps, the writing ability test, is typically found to correlate about .40 with FYGPA (Linn, 1975b, pp. 18-19). (In employment selections, correlations between a test and job performance of about .3 are common, and correlations seldom are much higher than about .6 (Linn, 1975a, pp. 45-46).)

The variance of the criterion may be divided into two parts--the part which is predictable from the predictor, and the part which cannot be predicted from the predictor. The variance of the criterion is equal to the sum of the variance of the predicted criterion scores and the variance of the errors of the estimate. It can be shown that the square of the correlation coefficient is equal to the proportion of the criterion variance which is predictable from another.

In the cases which we are considering, the proportion of variance of a criterion that is predictable from a predictor is typically rather small. For example, in undergraduate selection, a typical correlation coefficient of about .58 would indicate that only 34% of the criterion (FYGPA) variance is predictable, and 66% is unpredictable. In law school selection, a typical correlation of about .40 would indicate that only 16% of the criterion variance is predictable, and 84% is unpredictable. (In employment selection, a typical correlation coefficient of about .3 would indicate that only 9% of the criterion (performance on the job) is predictable, and 81% is unpredictable.) As the American College Testing Assessment Program (1973) has recently written:

> An important implication of the results is that while the prediction of college academic accomplishment by high school grades and test scores is good when compared with other types of prediction in education and the social sciences, the predictors still account for only a modest portion of the variation which occurs in the criterion. (p. 150)

Selection Efficiency

Given these comments, one might question whether these predictors can be of any help in making selections. One of the most important findings by psychometricians is that, within a certain selection situation, a correlation of .3, or even lower, may indicate that the predictor would be quite useful. It has been shown that the usefulness of a predictor for selection depends upon several factors in addition to the magnitude of the correlation. In particular, its usefulness depends on the proportion of applicants who are to be selected, which is called the selection ratio, the proportion of all applicants who would be successful if selected, which is called the base rate, and the importance of increasing the average performance on the criterion. The tables of such relationships are commonly called the Taylor-Russell Tables.

Suppose, for example, if the predictor were ignored and people were selected at random, that half the people would be considered successes and half failures (i.e., the base rate is .5). If 80% of the applicants had to be selected (the selection ratio is .8) to fill the positions, then the predictor would have to have a correlation of .55 with the criterion to yield a success rate of .58 (i.e., the proportion of selected people who end up being successful), which is an improvement of .08 over the base rate. On the other hand, if the selection ratio were .2, then a correlation of only .15 between the predictor and the criterion would be needed to yield a success rate of .58 (Linn, 1975a, p. 45 and Breland, 1974, pp. 11, 13). If the base rate were .5 and the selection rate were .2, then a correlation of .35 would yield a success rate of .70, while a correlation of .40 would yield a success rate of .73. (The correlation values of .35 and .40 are typical of law school selection processes.) Thus, given a base of .5 and a small selection ratio, a predictor that has a relatively small correlation with the criterion can lead to substantial improvements in the success ratio (Anastasi, 1968, pp. 134-138). (Remember, in the Defunis case there were about 1,500 applicants for the 150 positions, while in the Bakke case, in 1974, there were about 3,700 applications for the 100 positions.)

The Taylor-Russell Tables, of course, are based on certain statistical assumptions which are, at best, only an approximation to the statistical distributions that may be encountered in practice. Thus, although the tables cannot be used to supply precise values, the general point is sound. When the number of applicants greatly exceeds the number of positions, and when the base rate of the people is of small to moderate size, a predictor which has a relatively small correlation with

the criterion can probably be used to substantially improve the efficiency of the selection process (Linn, 1975a, p. 50 and ACTAP, 1973, p. 222 and Ch. VII). Quite simply, if there are enough applicants with high scores, then one might be able to fill the positons with applicants who are very likely to succeed.

In summary, the point that was first made by Taylor and Russell explains why an institution of higher education would use a prediction equation which may seem to be of little use because of the relatively small value of the correlation between the predictor and the criterion. Given certain conditions involving the selection rates and the base rates, the predictor can probably be used to increase substantially the efficiency of the selection process (Linn, 1975b, p. 5 and ACTAP, 1973, pp. 222, 250).

Given that using the prediction equation typically results in high selection efficiency, what can be said about the fairness of the selection procedure? Before I address this question, let me relate some other, important facts.

Some Important Facts About Test Scores and Certain Groups

Empirical studies have repeatedly shown that ". . . children of the poor tend to score lower . . . on standardized ability and achievement tests than do children of the affluent" (Stanley, 1971). Also, many studies have shown that members of certain minority groups (e.g. Blacks, Chicanos, Puerto Ricans) tend to score lower on ability and achievement tests than do members of the white majority (J. S. Coleman, et al, 1966). These differences in average test scores for different socio-economic or ethnic groups have been well documented and should be agreed upon by all. Attempts to explain why these differences occur remain controversial.

The important question here is whether the test has unfairly brought about these differences, or whether these differences were brought about by other factors. For example, there might have been background unfairness because some members of a certain group were denied a fair chance to acquire the necessary qualifications (Linn, 1975b, p. 6; Breland, 1974, p. 15). This kind of unfairness in the selection process is independent of any unfairness which may be brought about by using the test or other predictors. Also, the disparities in educational achievement may have been brought about by differences in cultural values and motivation; under these conditions, one would expect differences in test scores if the tests were measuring (or predicting) educational achievement. It is unclear whether any kind of unfairness would be attributed to the test in such cases of cultural differences.

The Charges Leveled Against Tests

Many people, however, have placed part or all of the blame for the differences on the tests. For example, some have argued that tests are a reflection of the majority white culture and inappropriate for members of other cultural groups. Thus:

> Many minority group individuals view existing testing practices as further obstacles to greater opportunity and status in education and employment, or as a more subtle attempt to perpetuate the stereotype of racial and/or ethnic inferiority. (Spencer, 1972, pp. 12-13)

Tests are often viewed not only as "gatekeepers, opening and closing pathways of human opportunity" (Hofman and Doctor, 1972, p. 14), but as tools ". . . in the perpetuation of racism on all levels of social and economic intercourse" (Jackson, 1975, p. 88). Tests are thought to provide a rationalization for discrimination by documenting "the basic faults and inferiorities of those being or about to be exploited" (Yette,

1971, p. 80). It was such cases and interpretations of tests that led the Association of Black Psychologists to call for ". . . an immediate moratorium on all testing of Black people until more equitable tests are available" (Williams, 1972, p. 20). In the case of Defunis v. Odegaard it was argued that it is not ". . . irrational to conclude that today's educational tests and majority-oriented college programs do not predict law school performance as accurately for the minority as they do for the non-minority applicant" (Brief of the Lawyer's Committee for Civil Rights Under Law, Defunis v. Odegaard, p. 20). In order to offset the difficulties with tests, "Changes may be limited to weighing standardized test scores differently for minority candidates than for the generality of applicants. An American Indian student who has lived all his life on a reservation and has not had the opportunity to attend best schools may not score as well on the LSAT as his white contemporary even though he has the same promise of academic performance" (Brief of the President and Fellows of Harvard College, Defunis v. Odegaard, p. 41). In other words, the tests predict "accurately" for whites but not for minority members. In the Bakke case, California Supreme Court Justice Tobriner claimed in his dissenting opinion, ". . . that the 'objective' academic credentials (i.e., tests and previous grades) on which the school had largely relied in the past did not accurately predict such minority applicants' qualifications and did not provide an equitable basis for comparison with other applications" (Dissenting Opinion, Bakke v. The Regents of the University of California, p. 39). He concluded that the school was entitled to "adjust its standards to overcome any built-in bias (p. 39).

Tobriner also argued that the school was well advised to de-emphasize Medical College Admission Test (MCAT) scores and grade point averages for minorities because "such academic credentials bear no significant correlation to an individual's eventual achievement in the medical profession" (p. 40). At this point, Tobriner's criticism is wide of the mark. The MCAT and grade point average are used to predict first-year grade point average, a task which is done reasonably well. Secondly, if these predictors are not related to professional achievement, then they should be de-emphasized for all applicants, not just for minority applicants. Tobriner agrees, but says that we should start with the minorities (p. 41, see also ACTAP, pp. 176-177). If schools of medicine and of law can find no valid predictors of professional achievement, then they should be careful in using professional achievement as a predictive basis for selection, for such selection procedures might be based on irrelevant distinctions, and therefore they would be unfair.

Even the federal courts seem to share a similar view about tests. "Statistical evidence showing racial disparity has been acknowledged by this court and the lower federal courts as sufficient to establish a prima facie case of discrimination" (Brief of the State of Ohio, Defunis v. Odegaard, p. 14). Novick and Ellis (1977) provide an interesting discussion of some of the important legal cases which bear on the issue of test fairness, including the cases of Greggs v. Duke Power Co. and of Albermarle Paper Co. v. Moody.

One must be careful, however, to remember that what is legally just may not be what is morally just. After all, a legal statute may have been consciously designed to achieve a result that is unjust, such as the conservation of economic advantage for a favored class as an end in

itself, a legal arrangement designed to ensure that one group profit at the expense of another. In other words, the purpose of the law itself may be arbitrary and invidious (Feinberg, 1973, p. 105). One must approach the legal cases with a clear understanding of the general differences between legal and moral justice.

Current Uses and Alternatives

At this point, I would like to make a few remarks about the actual practice of using predictors in college admissions (see Lunneborg and Radford, 1966).

As Linn has recently written, the increased number of applicants to law school has put considerable pressure on the admissions procedures and has tended to make the already fairly selective admissions requirements increasingly selective. For example, in 1967-68, the LSAT was taken by about 51,000 persons, while in 1971-72, it was taken by about 122,000 persons. Over the same period of time, however, the number of first-year law students increased from about 26,000 in 1967-68 to 38,000 in 1971-72. Thus, there was an increase of about 140% in the number of people taking the LSAT, but only an increase of 47% in the number of positions. As a consequence of this greater selectivity, the average LSAT score and undergraduate grade point average (UGPA) of first-year law students has increased substantially at many law schools (Linn, 1975b, p. 1). It may well be that these changes have also taken place at other kinds of educational institutions.

Furthermore, one should bear in mind that many educational researchers feel that the alternatives to tests and past performance have greater inadequacies at the present time so that the use of these predictors is not likely to be abolished in the near future. If tests are not used as

aids in making these necessary decisions, then other devices such as the interview, letters of reference, and subjective hunch must be heavily weighted. The record of validity of such procedures is a poor one relative to that of psychological tests (Breland, 1974, pp. 4, 6; Linn, 1975a, pp. 110-111; Schmidt & Hunter, 1974, p. 7).

Conditions for Fairness: Test Bias and Fairness

My major purpose in this essay is to subject the views of various educational researchers and psychometricians to philosophical criticism. One might wonder whether philosophical research has anything to offer to the educational researcher and psychometrician. After all, educational researchers are concerned with fairly technical issues in statistical analysis. Surely, a philosopher has no special expertise in these matters. Philosophers should leave these technical issues to those who have the special expertise.

If the concept "test bias" has remained a technical term in psychometric theory, then perhaps the above argument would be sound. What has happened in recent discussions, however, is that people have implied that the concept "test bias" has some logical relationship to the moral concept of fairness. For example, an expert in the field has written that we are dealing with the "issue of what constitutes test bias, or, conversely, what is fair test use?" (Linn, 1975b, p. 4). The Academic College Testing Assessment Program, which conducted a large empirical study into these matters, also held that the "converse of bias is fairness." Throughout its chapter on "Bias in Uses of Assessment," the report used the words "unbiased" and "fair" interchangeably (ACTAP, 1973, pp. 229-30, 272-3). A recent article in a journal issue devoted to the issue of test bias was entitled "Equity in Selection--Where

Psychometrics and Political Philosopy Meet (Cronbach, 1976). Again, writers have claimed that "the lack of general agreement on what constitutes fair test use creates real discomfort for any who would like to have a well-specified set of procedures to follow in order to insure fairness" (Linn, 1975b, p. 40) and that the real difficulty lies in answering the question "What is fair?" (Breland, 1974, pp. 14, 20).

It seems that one would be unable to understand fully the nature of the discussion in the research journals unless one understood that the term "test bias" or "test fairness" has been taken to be related to the moral concept of fairness. Even if a technical meaning of the term were preserved, the distinction between the tecnnical term and the moral term is likely to be lost when these issues are discussed in less technical journals. Most readers will take it that the word "fair" is being used in its ordinary, moral sense (Flaugher, 1974). The writers also have no qualms about using these terms in discussions of recent Supreme Court cases, such as the Defunis case, cases which involve matters of political and social philosophy.

My purpose here is not to convict these writers of error. Because fairness is an important moral concept, one should pay close attention to any attempts which may provide insights into the nature of fairness in selection. Such help would be welcome. On the other hand, care should be taken to make sure that these attempts do not distort the concept. If the concept becomes distorted, then to apply the concept to selection procedures may be to express moral approval where, in fact, no moral approval is warranted.

In the work which follows, I shall criticize various views of educational researchers and psychometricians which attempt to provide

sufficient conditions for the fairness of a test in its role in the selection procedure. In particular, I shall be interested in determining whether the suggested conditions are sufficient or necessary conditions for fairness.

Some Caveats: The Principle of Equal Access

Many philosophers hold that social justice requires that everyone have equal access to certain kinds of social goods. For example, it is justifiably claimed that everyone should have, at least, some minimum standard of living, education, and security (Frankena, 1962, p. 21). In these situations, the principle of social justice is the principle of equal access.

Now there are many who would argue that there are numerous social inequalities which are not necessary or justifiable, and that such inequalities should be eliminated. There may be a few who would contend that all social inequalities are unnecessary and unjustifiable, and that they ought to be eliminated. Those who opt for the latter positions are appropriately termed radical egalitarians (Bedau, 1971, pp. 163-180).

Further, it may be justifiably argued that some inequalities are just, and that other inequalities are justified by moral principles other than social justice. For example, it may be that a modern society can function only if certain role differentiations are made. These inequalities might be justified on the grounds of promoting the general welfare. At any rate, not all arguments on behalf of a certain kind of equality are arguments based on fairness (or justice), although it is generally true that justice involves equality (Bedau, 1971, p. 171; Brandt, 1962, p. 12; Berlin, 1961).

Let me return to the case at hand. Someone might argue that fairness requires that everyone who wants to go to college should be given an equal chance in the selection process. This principle relies on the principle of equal access which, when it is applied under conditions of scarcity, becomes the principle of equal chance. Given the number of positions available, P, and the larger number of applicants, A, every applicant should have the chance P/A of being selected.

It is difficult to assess here whether in this kind of a situation fairness requires that everyone be given an equal chance. The argument will be different from the arguments which concluded that everyone was entitled to minimum amounts of certain goods necessary for a good life. In fact, one moral philosopher, John Rawls, holds that it would be just to distribute advantageous offices and position on the basis of equal opportunity for the best qualified (Rawls, 1971, pp. 83-90). Furthermore, the argument must address itself as to whether an open-admissions policy would undermine the institution's purposes in providing a quality or a particular kind of education. I am not sure that open admissions need have such a corrosive effect, but certainly a proponent of the principle of equal chance must show that such things are not problematic.

Clearly, except for those situations which have open-admissions selection procedures that give an equal chance to all applicants, all institutions of higher education violate the principle of equal chance because they select their students on the basis of some set of qualifications. If fairness were to require equal chances, then the selection procedures of most institutions would be unfair.

It seems to me that fairness does not require equal chance, although I can't argue for it here. Indeed, it is more plausible that fairness

in selection procedures is related to two other principles, the principle of equal opportunity for those suitably qualified, and the principle of equal accuracy. In the criticism which follows, these latter principles will be employed. The issue of test fairness arises only where justice requires something other than the principle of equal chance.

Various Policies and Procedural Fairness

Suppose an institution were to define its set of qualifications and requirements and to develop a procedure for selecting students according to these requirements. No matter what the qualifications were or the kinds of applicants which were treated in various ways, if an administrator were to follow the procedure properly, he would be serving procedural fairness. Analogously, whatever the content of the law and the nature of the treatment to be given out, if a judge were to follow the law impartially (and properly), he would be serving procedural fairness. If the institutions were to specify adequately their selection procedures, then their following of these procedures would serve procedural fairness.

Thus, each of the conditions for fairness which educational researchers have put forward could be satisfied by a given policy and procedure so that each and every policy could serve procedural fairness, even though the policies would differ greatly in the kinds of treatment they give to certain sorts of people. While almost every selection procedure would violate the principle of equal chance, almost every selection procedure could be procedurally fair.

Procedural fairness, however, has moral value only insofar as it serves or secures substantive fairness. I shall argue that, even if a policy satisfies one of the purported conditions of educational researchers, it does not follow that it serves substantive fairness.

Principles of Nondiscrimination

There are other principles which are independent of the principles of fairness and which supposedly have moral force. The ideal of nondiscrimination holds that there is something degrading or morally offensive in treating differently people who differ in certain ways (e.g., in the color of their skins). The related principle of nondiscrimination holds that, as a matter of fact, certain kinds of injustices (e.g., racial injustice) can best be fought and eliminated by making it a rule, without exception, that using certain characteristics (e.g., racial characteristics) as the basis for distributing certain goods should be morally prohibited. Hunter and Schmidt have talked about such principles under the label "qualified individualism" (1976, p. 1054).

Given that these principles are morally valid (or sound), then any procedure which uses racial characteristics, say, in order to insure higher levels of predictive accuracy to achieve equal opportunity, or to implement certain kinds of group quotas, would be in violation of these principles. The principles of nondiscrimination would be violated by trying to integrate (or segregate) a student body because one must use racial characteristics to bring about the integration. Of course, whether these are valid (or sound) moral principles and how morally important are these principles are questions which must be answered before we can understand the kind and degree of obligation which one has to conform to them. A. I. Melden has recently argued that programs of action which select certain groups as their peculiar beneficiaries are sometimes required on the grounds of human rights and justice (Melden, 1977, Chapter VII, p. 253). Apparently, then, the principle of nondiscrimination can be overridden by other moral principles.

Educational Researchers' Conditions for Fairness

1. Equal Test Score Averages: The Position Stated

A selection procedure (which uses a test to select its students) is fair if, and only if, the groups (e.g., the majority and the minority) have the same average scores on the selection test. This position is implicit in many of the viewpoints held by those who have charged that tests are unfair to certain minority groups.

Criticism. The position is most plausible if it is assumed that the groups do not differ on the characteristics which the test purports to measure (or predict)--that is, the criterion measure. For then, it seems reasonable to conclude that differences in average test scores for the groups are brought about by the imperfections or the unfairness of the test. But, even if the groups do not differ on the criterion measure, the fact that there are equal test score averages is still not a sufficient condition for fairness. First, it may turn out that one group has a wider spread of test scores than the other group, and that if the institution picks those applicants with the higher test scores, then few of those from the other group (who are as equally qualified as many of those selected) will be admitted. Second, it could turn out that the rank-ordering in one of the groups is distorted (or even reversed) and, therefore, that if one picks those applicants with the higher test scores, few of those from that group (who rank as high on the criterion as those who are selected) will be admitted. Those remarks show, consequently, that even if the groups have the same score averages and the same criterion scores, other conditions must be satisfied before use of the test is fair.[2]

If the groups were different on the criterion scores, then it appears reasonable to expect that the groups would be different on the test scores. For the test might reflect the rank ordering of the total population on the criterion measure, even though the average test scores for the groups are different. Under these conditions, selecting those with the highest test scores would result in a fair selection. Therefore, a fair test does not require that the groups have the same average test scores; that there are differences in average test scores does not imply that the test is unfair or biased.

In conclusion, equal test score averages are neither a necessary nor a sufficient condition for the fairness of a selection procedure which uses a test. Psychometricians, too, have long recognized the inadequacy of this condition, although the general public still seems to regard it as a necessary or a sufficient condition.

2. Equality of Prediction Systems (The Regression Condition)

The position stated. Suppose that the selection procedure relies heavily on the prediction equation. A test is fair if it is unbiased; a test is unfair if it is biased. Here is the general description of test bias: "In the psychometric sense, test bias refers to overprediction or underprediction of criterion measures. If a test consistently underpredicts criterion performance for a given group it shows unfair discrimination or 'bias' against the group" (Anastasi, 1968, p. 559). Since the most frequently used prediction strategy is linear regression, one gets the more specific description of test bias: "A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the

test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup" (Cleary, 1968, p. 115).

This last definition, or some minor variant of it, has been the most widely accepted definition of test bias within a predictive context for the last 20 years (Darlington, 1971, p. 73 and Linn, 1975b, p. 7). I shall refer to it as the regression condition. Furthermore, "these definitions assume that fairness is achieved by selecting on the basis of predictions of a criterion score and lead to the examination of regression equations in the separate groups for consistent errors of prediction" (ACTAP, 1973, p. 231). A test is unbiased in this sense, then, when the regression lines (or the prediction equations) are the same; that is, the regression lines for each group have the same slope and the same criterion axis intercept. If the regression lines (or the prediction equations) are not the same, then the separate regression equations should be used in order to be fair[3] (ACTAP, 1973, pp. 231, 272). (See Figures 1a, 1b, 2a.)

Insert Figures 1a, 1b about here

In particular, suppose the slopes of the regression lines are not the same for the two groups and that the regression line of one group say group 1, is used to predict the performance of both groups. Since colleges usually select those who have the highest predicted scores first, using the test in this selection procedure is likely to overpredict systematically and consistently members of group 2. The procedure is biased in favor of group 2 because it overpredicts systematically the

performance of members of group 2 and thereby gives them an undeserved advantage in the selection process (see Figure 1a).

Suppose the slopes of the regression lines are the same, but that the criterion-intercepts are not the same. Again, suppose that the regression line of one group (say, group 2) is used to predict the performance of both groups. Since colleges usually select those who have the highest predicted scores first, using the test in this selection procedure is likely to systematically underpredict (and consistently so) the members of group 1. The procedure is biased against group 1 because it underpredicts systematically the performance of members of group 1 and thereby puts them at an undeserved disadvantage in the selection process (see Figure 1b).

After studying the empirical evidence, the ACTAP report concludes that "to be fair to racial-ethnic groups according to the regression model, separate within-group regression equations should be used to select students with the highest predicted GPA's regardless of group membership. The use of a single combined equation is unfair to the majority group by overpredicting minority GPA's" (1973, p. 272). Whether ACT has reached a sound conclusion is something I shall leave to the side. What is important here is that it states a logical condition for fairness.

Criticism. The concept of test bias, then, was originally intended to refer to a particular aspect of the prediction system: consistent, systematic nonzero errors of prediction which result for members of a certain subgroup. It seems to me that this is an important and unique kind of unfairness which may result from using a certain test in a selection procedure. Still, that a test is unbiased in this sense is,

at best, a necessary condition for a selection system to be fair. Without considering in detail different kinds of prediction strategies, it is hard to conclude whether a test's being unbiased is a necessary conditon. It seems to me that it is, but I know of no general way to show that this is so.

However, by considering the following example, it can be seen, however, that the condition is not a sufficient condition for substantive fairness. Suppose that each group has identical criterion scores, test scores, and regression equations. In this situation, the scatter plots would coincide and the test would be unbiased according to the psychometrician.

Suppose there are N openings at the institution. The admissions officer is interested in admitting those applicants who will probably achieve the highest criterion scores and, in order to do so, he selects applicants who have the highest test scores. He selects those with the highest test scores until the positions are filled. Thus, the number of applicants who are selected equals A + D = N (use the group 1 diagram in Figure 3a). Although the selected procedure is "unbiased" (i.e., it satisfies the regression condition), the selection procedure has not achieved fairness. Substantive fairness requires that each of the highly (or suitably) qualified applicants be given an equal opportunity to obtain one of the N positions. But, because of the test's lack of discrimination between As and Ds, the number of positions effectively open to the highly qualified is reduced to N - D. (Here, the test is unfair because its use results in treating unequals equally.) Further, because the test illegitimately discriminates between As and Bs, those applicants who are highly qualified are denied equal opportunity. Given

the typical predictive validities with moderate or low selection ratios, B is likely to be the same size or two to three times larger than A, depending on level of the criterion cut point. (Remember that the typical predicted validities are .58 in undergraduate selection, .40 in graduate selection, and .30 in general employment selection.) Those in B are not only denied equal opportunity, but they are summarily rejected by the test. Those in A are also treated unfairly. They have no claim to the position, but only a claim to an equal opportunity to obtain the position. Here, the test is unfair because its use results in treating equals unequally. Thus, the selection procedure, whatever may be its value as an effective selection device of qualified students, does not achieve fairness--in fact, it is unfair to members of both the minority and the majority groups.

Insert Figures 2a, 2b about here

In conclusion, I have argued that the regression condition (i.e., equality of prediction equations) is not a sufficient condition for fairness. In fact, although the regression condition is satisfied, current selection procedures grossly violate the principle of fairness--the principle of equal opportunity for those suitably qualified.

The Equal Risk Condition (Einhorn and Bass)

The position stated. The following descriptions have been given and taken to be compatible. "Unfair test discrimination exists when persons with equal probabilities of success on the job have unequal probabilities of being hired" (Guion, 1966, p. 26). "Einhorn and Bass (1971) proposed a model for test bias which takes into account, for each

subpopulation, the probability of success associated with the applicant's predicted criterion score as suggested by the regression" (Petersen and Novick, 1976, pp. 18-19). The best description, however, is the following: "A selection rule for two groups would be considered fair according to the Equal Risk Model if the probability that a person from group 1, with a test score equal to group 1's cut score will be above the minimum level of acceptable performance on the criterion" (Linn, 1975a, p. 141). (See Figure 2b.) This later description states in precise terms the way that the condition has been applied in actual situations. The ACT program report of 1974 concurs with this description.

Equal risk can be seen as a refinement of the regression equation strategy. The institution may not want to take people for whom the probability of success is 50%, for example, but may wish to accept those whose chance of reaching the specific level of performance is, say, 80%. As Bass has written recently, ". . . all those selected (with scores at or above the cutoff score), regardless of group membership, will be within the limits of the maximum risk of failure which the institution is willing to accept" (1976, p. 611). In other words, by setting the same minimum probability Z of achieving a certain level of success, given the test score, the institution is likely to admit applicants so that Z percent or more of these will achieve at the certain level of criterion success. The institution sets the levels of risk which it is willing to take.

The advantages and disadvantages of the Equal Risk Strategy are similar to those of the regression model, except that the condition is sensitive to group differences in the standard errors of the estimate (Linn, 1975a, p. 153).

It should be clear that the primary virtue of this selection strategy is that it probably ensures a group of students of which Z percent or more will achieve at least the minimum acceptable level of criterion performance by setting the appropriate cutting-off points for each group. In other words, given that one can legitimately establish the minimum (or suitable) level of criterion performance, this strategy is efficient, especially when the selection ratio is low.

Granted, then, that under certain conditions which are approximated in actual practice, the selection procedure is efficient. But is it fair? No. The criticism of the equal risk condition is identical to the criticism of the regression condition--it violates the principle of equal opportunity for those who are suitably qualified. I omit the argument for brevity. Thus, the equal risk condition is not a sufficient condition for fairness.

4. The Thorndike Constant Ratio Condition

The position stated. Thorndike described the position in the following way. If the conditions set forth by the regression strategy are satisfied, it can be shown "that when the two groups differ appreciably in mean test score, the regression procedure which is 'fair' to individual members of the group scoring lower on the test, is 'unfair' to the lower group as a whole in the sense that the proportion qualified on the test will be smaller, relative to the higher scoring group, than the proportion that will reach any specified level of criterion performance. . . . The qualifying scores on a test for each group should be set at levels that will qualify applicants in the two groups in proportion to the fraction of the two groups reaching a specified level of criterion performance" (Thorndike, 1971, p. 63).

In other words, the selection procedure using the test is fair if the proportion of those selected from group 1, divided by the proportion of those qualified from group 2, varies directly as the proportion of those qualified from group 1, divided by the proportion of those qualified from group 2. Again, using the test is fair if the proportion of those selected from group 1, divided by the proportion of those qualified from group 1, equals the proportion of those selected from group 2, divided by the proportion of those qualified from group 2. (That is, $(A_1 + D_1) / (A_1 + B_1) = (A_2 + D_2) / (A_2 + B_2)$.) I shall refer to this as Thorndike's condition. It is claimed that this is a sufficient condition for fairness (see Figure 3a).

Insert Figures 3a, 3b

Criticism. Thorndike's suggested condition seems plausible because it is easily confused with the following case. Suppose that there are N places at the institution and that there are more highly qualified applicants, M, than places. If each of the suitably qualified applicants were given an equal opportunity to obtain the position, then he would be given the chance to obtain the position equal to N divided by M. If there were M_1 suitably qualified people in group 1, then the M_1 of them should have the chance N/M; if there were M_2 suitably qualified people in group 2, then the M_2 of them should have the chance N/M. Thus, the expected number of suitably qualified people selected from group 1 is M_1N/MT and from group 2 is M_2N/M. The proportion of suitably qualified people selected from group 1 is M_1N/MT_1 and from group 2 is M_2N/MT_2, while the proportion of suitably qualified people in group 1 is M_1/T_1

and in group 2 is M_2/T_2, where T_1 (T_2) is the total number of applicants from group 1 (2). It follows, then, under these circumstances, that Thorndike's condition would be satisfied. It is important to notice, however, that it has only been shown that the condition is necessary, under special circumstances, according to the principle of equal opportunity for those suitably qualified.

It can be shown, however, that the condition is not sufficient. Consider the data for selecting students by predicting their criterion performance on the basis of test scores, and suppose that Thorndike's condition is satisfied. The number of positions to be filled is N. Therefore, $A_1 + D_1 + A_2 + D_2 = N$. Notice that $D_1 + D_2$ of the possible openings will not be allocated to the suitably qualified because the test's lack of discrimination results in its not being able to distinguish between A_1s and D_1s (and between A_2s and D_2s). Thus, the number of positions for which the suitably qualified might compete is already reduced from N to $N - (D_1 + D_2) = A_1 + A_2$. Furthermore, because the test illegitimately discriminates between the suitably qualified in A_1 and B_1 (and in A_2 and B_2), those in B have no chance at being selected. (In actual cases B may well be the same size or two or three times larger than A.) What fairness requires is that every suitably qualified person be given an equal opportunity to obtain the position. Not only is it likely that a majority of the suitably qualified is being denied equal opportunity, but a majority is summarily rejected. Even those members of A_1 (and A_2) were given unfair treatment. They have no claim in the name of fairness to the position, but only a claim to an equal opportunity to obtain the position.

In summary, I have argued that Thorndike's condition seems plausible because it states a necessary condition for fair selection of suitably qualified students, under certain circumstances, according to the principle of equal opportunity for the suitably qualified. But the facts about current selection procedures show that the actual situations are different from those circumstances in which the condition is a necessary one. At any rate, in these situations, it can be shown that the condition is not sufficient. In other words, given the certain facts about selection procedures, the condition is necessary but not sufficient. Since the suitably qualified applicants (of both the majority and minority groups) are denied equal opportunity to obtain the positions, current procedures are fair in only the grossest sense.

5. The Cole Conditional Probability Condition

The position stated. Cole described the positon in the following way: "The basic principle of the conditional probability selection model is that for both majority and minority groups whose members can achieve a satisfactory criterion score (Y Y*) there should be the same probability of selection regardless of group membership" (1973, p. 240). From this description, it seems that one can determine whether the procedure is fair by looking at group characteristics. In the following description of Petersen and Novick, it seems that one is talking about the characteristics of individuals. "All applicants, regardless of group membership, who, if selected, are capable of being successful should be guaranteed an equal, or fair, opportunity to be selected" (Petersen & Novick, 1976, p. 10). Linn has presented the following description. "A selection procedure would be fair . . . if the probability of being selected, given satisfactory performance, is the same

for all groups" (Linn, 1975b, p. 38). The description given by Linn is the best because it states clearly what has come to be the accepted operational definition of the condition. In particular, Cole's condition is satisfied when the proportion of those selected and qualified from group 1, divided by the proportion of those selected and qualified from group 2, varies directly as the proportion of those qualified from group 1 is divided by the proportion of those qualified from group 2. (That is, $A / (A_1 + B_1) = A_2 / (A_2 + B_2)$. It is claimed that this is a sufficient condition for fairness (see Figure 3b).

Criticism. Cole's suggested condition seems plausible because it is easily confused with the following case. Suppose that there were originally N places available, but, due to unfortunate, external difficulties, fewer than N places are available, N_1. If each of the suitably qualified applicants were given an equal opportunity to obtain the position (under scarcity), then he would be given the chance of obtaining the position equal to N_1 divided by M, where M is the total number of suitably qualified applicants. If there were M_1 suitably qualified people in group 1, then the M_1 of them would have the chance N_1 divided by M; if there were M_2 suitably qualified people in group 2, then the M_2 of them would have the chance N_1/M. Thus, the proportion of qualified people selected from group 1 is M_1N_1 divided by MT_1 and, from group 2, M_2N_1 divided by MT_2. The proportion of qualified people in group 1 is M_1 divided by T_1 and, in group 2, is M_2 divided by T_2. It follows then, that Cole's condition would be satisfied. It is important to notice, however, that it has only been shown in this sample that the condition is a necessary one according to the principle of equal opportunity for the suitably qualified in special situations.

It can be shown that in general the condition is neither necessary nor sufficient. Again, consider the actual data for selecting students by predicting their criterion performance on the basis of test scores and suppose the number of positions to be filled is N. Therefore, $A_1 + D_1 + A_2 + D_2 = N$. Again, notice that $D_1 + D_2$ of the possible openings will not be open to the suitably qualified, and that Cole urges that, in evaluating the fairness of the test, one should overlook this. But it is because of the test's lack of discrimination between A_1s and D_1s (and between A_2s and D_2s) that this unfortunate reduction occurs in the number of places available for the suitably qualified. The test itself has brought about this reduction and should be judged accordingly. A necessary condition, according to the principle of equal opportunity for the suitably qualified, is the Thorndike condition, not the Cole condition.

Furthermore, because the test illegitimately discriminates between the suitably qualified in A_1 and B_1 (and A_2 and B_2), those in B have no chance at being selected while those in A have probability 1 of being selected. (Again, it should be noted that B may be the same size or two to three times larger than A.) What fairness requires is that each suitably qualified applicant be given the same opportunity of being selected. Thus, Cole's suggested condition is not a sufficient condition.

It is interesting to notice that the condition suggested by Cole can be shown to be related to the derivative principle of fairness. The position is as follows. It is argued that perfect accuracy is an unrealistic goal and that errors in prediction will be made. Given that errors will be made, fairness requires that each person be predicted to the same degree of accuracy. If a person is suitably qualified, then it

should be that he is predicted with probability P to be suitably qualified; if a person is not suitably qualified, then it should be that he is predicted with probability P to be not suitably qualified. Although mistakes in prediction will be made, every person has the same risk of being improperly classified. The risk is distributed equally to all individuals. Suppose, also, that there are as many positions, N, as there are people who are predicted to be suitably qualified. Under these conditions, it follows that the proportion of those qualified and selected from group 1 is A_1 / T_1, while the proportion of those qualified from group 1 is $(A_1 + B_1) / T_1$; the proportion of those qualified and selected from group 2 is A_2 / T_2, while the proportion of those qualified from group 2 is $(A_2 + B_2) / T_2$. Since $A_1 / (A_1 + B_1) = A_2 / (A_2 + B_2) = P$, Cole's condition is necessarily satisfied under these conditions.

So far, I have shown that under special conditions, Cole's condition is a necessary condition for fairness, where the principle of fairness is the principle of equal accuracy. The condition, however, is not a sufficient condition. For the condition may be satisfied when one looks at group characteristics, but it need not follow that each person is predicted with the same accuracy. For example, members of one-half of a group may be predicted with the accuracy .5 P, while members of the other half of the group may be predicted with the accuracy of 1.5 P. Thus, Cole's condition would be satisfied (for the group), but the principle of equal accuracy would be violated for the individuals. (The reader can create more extreme cases of inequality for himself.)

One further comment about the principle of equal accuracy. It seems that the value of P, that is, the degree of accuracy, should be rather large, surely larger than one half. For, if the degree of accuracy

were .001 but the same for all, I would hesitate to say that it would be even a derivative principle of fairness. Therefore, it seems that the principle of equal accuracy is required, but that the principle includes the condition that the degree of accuracy be at a reasonable level. Although there is a gap between achieving the substantive purpose (i.e., selecting those who are suitably qualified and rejecting the others) and achieving the procedural purpose (i.e., selecting those who are predicted to be suitably qualified), the relationship between a person's qualifications and his predicted qualifications is reasonably tight, and it is the same for all individuals. Therefore, the principle of equal accuracy is, indeed, a principle of substantive fairness.

In conclusion, I have contended that Cole's condition is neither a necessary nor a sufficient condition for fairness according to the principle of equal opportunity for those highly (or suitably) qualified. Also, I have argued that Cole's condition is, in certain situations, a necessary condition but not a sufficient condition according to a derivative principle of fairness--the principle of equal accuracy.

Additional Comments

(1) Petersen and Novick offered a criticism of the Thorndike condition and the Cole condition. For brevity, I will consider the criticism they presented of Thorndike's condition, which they called the Constant Ratio Model. According to the Constant Model, the test is fair if the ratio $(A + D) / (A + B)$ is the same for both the majority and the minority groups, for example. Petersen and Novick argued that similar logic should lead one to require that the converse ratio $(C + B) / (C + D)$ be the same for both groups in order to achieve fairness. Roughly, if those who are suitably qualified should be given an equal opportunity to

be accepted, then those who are not suitably qualified should be given an equal chance to be rejected (Petersen & Novick, 1976, p. 21).

What is interesting is the following. They stated that as a matter of fact, a selection procedure (which uses the criterion-test relationship) cannot satisfy both of these conditions, but that a "model" must take both sides (selection-success and rejection-failure) of the test bias issue into account. Since the Thorndike model is "contradictory" because both ratios cannot be satisfied at once, Petersen and Novick imply that this is a conclusive reason for rejecting the Thorndike model, and for preferring another "definition" of fairness.

This argument is fallacious as the following example will show. Legal justice requires that the innocent be found innocent and the guilty be found guilty. As a matter of fact, however, as one does better in achieving the task of finding the innocent to be innocent one does worse in achieving the task of finding the guilty to be guilty, and vice versa. First, suppose these goals are of equal moral value. Does it follow that if only one of them can be achieved, then both must be rejected? It does not. It seems perfectly reasonable and moral to choose to achieve one or the other. Each of these goals may have considerably more moral weight than other goals. Second, it might turn out that one goal is morally worth more than the other, although the "logic" is similar for valuing both of them. In Western culture, it is believed to be worse to convict an innocent man than to free a guilty man, although the cases seem logically similar. Analogously, it might be that we are more concerned about securing equal opportunity for those who are suitably qualified than about securing an equal chance to be rejected for those who are not suitably qualified.

In conclusion, Petersen and Novick must provide additional premises before they can conclude that the Thorndike condition should be rejected.

(2) It is interesting to note that a recent study has concluded that current selection procedures which use the criterion-test relationship and which select the students with highest predicted scores do not come close to satisfying either the Thorndike condition or the Cole condition, while they do come quite close to satisfying the regression condition and the equal risk condition. The report concluded that to satisfy the Thorndike condition (or the Cole condition), selection cutting points, or predicted grade point average (i.e., the predicted criterion) (from either separate or combined equations but, preferably, separate ones) should be lower for minority students than majority students, resulting in selection of substantially more minority students than a procedure which satisfies the regression condition (or the equal risk condition (ACTAP, 1973, p. 239-253, 272)). The report also pointed out that a policy which satisfies the Thorndike or the Cole condition will result in lower expected success rates among those selected. That is, the selection process will be less efficient (ACTAP, p. 252). By now, you should have a good idea why this is so.

(3) Some people have argued for certain selection policies on grounds other than those of fairness. For example, those who are interested in quotas and those who are interested in integration, without regard to qualifications, would not be fully satisfied with a policy which satisfied the Thorndike (or the Cole) condition because the policy would take the proportion of suitably qualified members in the groups to be a relevant factor. Those interested in quotas feel the proportion of those selected from each group should vary as the proportion of each

group's population in the city (state, or nation), regardless of the number qualified and of the characteristics of those who apply.

Those who are interested in goals and those who are interested in integration, which are intended to admit only qualified members of each group, would probably accept as a practical compromise a policy which satisfies the Thorndike (or Cole) condition if they believe that the proportion of qualified applicants in the groups is a legitimate reference. Further, if they believe that the legitimate reference is the proportion of qualified applicants in each group's population in a city (or state) then it is likely they would not be fully satisfied with either the Thorndike or Cole policies because these policies would admit fewer minority applicants than would a policy based on population figures. (Remember, it is likely that minority groups have relatively lower base rates of criterion achievement, that they have relatively lower rates of application, and that they have a smaller group size.)

Those who are interested in quotas, those who are interested in goals, and those who are interested in integration, however, might decide in favor of a policy which satisfies the Thorndike or Cole condition after they have compared the consequences of a policy which satisfies the regression (or equal risk or equal probability) condition. The Thorndike (or the Cole) policy would demand that proportionately more minority applicants and proportionately less majority applicants be selected than is the case under the current regression-based policies. (See also Breland & Ironson, 1976.)

Whatever the merits of the Thorndike condition, my arguments show that it is not a sufficient condition for fairness. A justification must be provided on grounds other than those of fairness.

6. Equal Probability

The position stated. The following descriptions have been given. In an earlier description it seemed that the condition was concerned with individuals. "All applicants, regardless of group membership, who are selected should be guaranteed an equal, or fair chance of being successful" (Petersen & Novick, 1974, p. 19). In a later description, it seemed that the condition was concerned with group characteristics. "According to the Equal Probability Model, within the selected group from each sub-population . . . , the proportion of successful performers should be the same" (Petersen & Novick, 1976, p. 13). This later description agrees with a recent description by Linn, that is, "The Equal Probability Model is based on the notion that the probability of success (or the risk of failure) of all selected applicants from one group should be the same as that from any other group" (Linn, 1975a, p. 154). Although these descriptions differ in various ways, the writers agree in operationalizing the condition in the following way. The equal probability condition is satisfied when the proportion of those qualified and selected from group 1, divided by the proportion of those qualified and selected from group 2, varies directly as the proportion of those selected from group 1, divided by the proportion of those selected from group 2 (that is, $A_1 / (A_1 + D_1) = A_2 / (A_2 + D_2)$. It is claimed that if the equal probability condition is satisfied, then the selection procedure is fair. That is, this condition is a sufficient condition for fairness.

Criticism. Perhaps there are good reasons for arguing that a selection procedure should satisfy this condition. For example, such a selection procedure might be quite efficient. In terms of fairness, however, this position seems plausible because it is easily confused

with the following case. Suppose that the test is such that each and every person has the same degree of evidence. That is, every person is either predicted with probability P to be suitably qualified, or predicted with probability P to be not suitably qualified. Suppose, also, that the number of those who are predicted with probability P to be suitably qualified is equal to the number of positions N, so that all of them will be selected. Therefore, the proportion of those qualified and selected from group 1 is equal to A_1/T_1, and the proportion of those selected from group 1 is equal to $(A_1 + D_1) / T_1$. The respective proportions for group 2 are A_2/T_2 and $(A_2 + D_2) / T_2$. Since $A_1 / (A_1 + D_1)$ equals P, and $A_2 / (A_2 + D_2)$ equals P, it follows that the equal probability condition is satisfied. Given that every person has the same degree of evidence, then it necessarily follows that the equal probability condition is satisfied.

Notice, however, that current selection procedures which use the criterion-test relationship do not satisfy the equal evidence condition. It is just not the case that every individual either is predicted with probability P to be highly (or suitably) qualified, or is predicted with probability P to be not highly qualified. There is no reason to expect that the equal evidence condition will be satisfied in the near future.

Furthermore, this principle of equal evidence is logically distinct from the principle of equal opportunity for those suitably qualified and the principle of equal accuracy. The principle of equal accuracy holds that, given that a person is suitably qualified (or not suitably qualified), then he should be predicted with probability P to be suitably qualified (or not suitably qualified). The principle of equal evidence holds that, given that a person is suitably qualified (or not suitably

qualified), then he should either be predicted with probability P to be suitably qualified or be predicted with probability P to be not suitably qualified. If a selection procedure were to satisfy the principle of equal accuracy, then it would satisfy the pinciple of equal evidence; but, if it were to satisfy the principle of equal evidence then it would not follow that it satisfies the principle of equal accuracy. Indeed, it seems empirically possible that a selection procedure might satisfy the principle of equal evidence and yet a large proportion of the suitably qualified might be predicted to be not suitably qualified.

The principle of equal opportunity for those suitably qualified holds that each suitably qualified applicant should have the same chance of obtaining one of the positions. A selection procedure which satisfies the principle of equal evidence, however, could reduce the number of available positions because it cannot precisely discriminate between As and Ds. Furthermore, the procedure could unfairly discriminate between As and Bs so that the Bs are summarily rejected while the As are accepted.

When the selection ratios are moderately low (say, less than 50%) and there is some correlation between the predictor and the criterion (say, .3-.5), then a selection procedure can be highly efficient; the proportion of those qualified and selected, divided by the proportion of those selected is likely to be high (say, .7 or higher). These selection procedures, however, are likely to be unfair either because they do not fairly distribute opportunity to those suitably qualified or because they do not fairly distribute the risk of misclassification.

Finally, the principle of equal evidence is, at best, a <u>procedural</u> principle; it is not a <u>substantive</u> principle of fairness. One might believe it is a substantive principle because one has failed to see the

distinction between procedural fairness and substantive fairness. It is true that the best way to ensure substantive fairness is to collect empirical evidence about the applicant's characteristics in order to decide whether he meets the qualifications. This way is empirically preferable to intuition, to hunches, to guessing, and so on. But the value of following this way depends upon the substantive fairness which results from following the procedure. Likewise, the value of the procedural principle of equal evidence depends upon whether substantive fairness is achieved. As I have argued above, the principle of equal evidence may be satisfied and yet neither the principle of equal chance for those suitably qualified nor the principle of equal accuracy need be satisfied.

As one might imagine, it can turn out that the principle of equal evidence is satisfied and that the (aggregate) efficiency of the selection procedure is quite high. The principle is not concerned with how and to which individuals the opportunities (or the risks of misclassification) are distributed within the group of applicants. But fairness requires more than that the degree of evidence be the same for all applicants. There has to be some connection with the accuracy of prediction in the individual case as well as the degree of evidence in order for substantive fairness to be served. A procedure which satisfies the principle of equal evidence in typical cases would predict some arbitrarily defined subgroup of the highly qualified to be suitably qualified while it would predict others of this group to be not suitably qualified. A similar thing would happen for the group of those not suitably qualified--an arbitrarily determined subgroup would be predicted to be not suitably qualifed while others of this group would be predicted to be suitably

qualified. Here, the lack of accuracy in individual cases balances out on the average so that selecting those who are predicted to be (probably) suitably qualified results in an efficient selection procedure.

This line of reason also shows that it is mistaken to say "it is fair to each individual, since each is accepted or rejected on the best possible estimate of his future performance" (Hunter & Schmidt, 1976, p. 1064). Such a statement not only blurs the distinction between procedural and substantive fairness, but it even violates the essentially comparative aspect of fairness.

It seems that the principle of equal evidence lies on the moral scale somewhere between the formal principle of justice (i.e., treat similar cases similarly and different cases differently, regardless of the make-up of the classes of the people, and of the kinds of treatment) and the substantive principles of fairness (e.g., the principle of equal chance for those suitably qualified and the principle of equal consideration).

In conclusion, I have argued that the principle of equal evidence and the condition of equal probability do not insure substantive fairness. If there is to be a justification for the current policies based upon these conditions, then it will probably be in terms of aggregate efficiency. As a matter of fact, the current use of tests to predict criterion performance is likely to be highly efficient but unfair in only the grossest sense.

Culture-Modified Criterion

The position stated. Darlington has written that, "There is often a conflict between the two goals of low cultural discrimination and high validity, and no simple mechanical resolution of this conflict is gener-

ally possible, since an optimum balance between the two goals necessarily depends on a subjective judgment concerning their relative importance" (1971, p. 71). In other words, there is a conflict between fairness and selection efficiency. Darlington suggested that, instead of predicting the criterion variable Y, a variable (Y - KC) be constructed where K is determined by a subjective value judgment on the part of the decision maker. He urged that, "The term 'cultural fairness' be replaced in public discusssions by the concept of 'cultural optimality.' The question of whether a test is culturally optimum can be divided in two: a subjective, policy-level question concerning the optimum balance between criterion performance and cultural factors (operationalized as the optimum value of K), and a purely empirical question concerning the test's correlation with the culture-modified variable (Y - KC) and where the correlation can be raised" (pp. 77-78).

Each institution, then, must first choose a value of K, indicating whether there is special value in the selection of members from some subpopulation. That is, the decision maker must answer the question, "How many units on Y are considered equivalent in value to one unit on C?" (Petersen & Novick, 1976, p. 23). Thus, group membership (or culture) is also part of the criterion.

Criticism. Whatever may be the merits of this strategy in finding a "culturally optimum" solution, in general, the culture-modified criterion strategy provides neither a necessary nor a sufficient condition for selection procedure fairness. In special situations, however, it does provide a necessary condition. Suppose that compensatory justice does require that a certain group of people be treated as they would have been if they had not suffered unfair disadvantage. Thus, given

their actual predicted criterion scores, an amount equal to K is added to them to bring them up to what they would have been. Therefore, the selection procedure uses the empirical realtionship between the tests and the variable (Y - KC), where C is equal to 1 if the applicant is a member of the unfairly disadvantaged group, and otherwise, C is equal to 0. Here, the value of K might be the difference between the average scores of the unfairly disadvantaged group and the average scores of the others. Although such a policy may be required by justice, it is likely to lower selection efficiency.

The strategy is not sufficient, however, for it is built upon the criterion-test relationship which is used in the regression strategy. The strategy is subject to the same limitations; it is unfair. Both the principle of equal opportunity for those suitably qualified and the principle of equal accuracy are violated. I omit repeating the arguments. (For further criticism of this method and Darlington's Graphic Method, see Linn, 1975a, pp. 186-187, 188-193.)

The Decision-Theoretic (or Equal Expected Utility) Approach

The position stated. So far, I have discussed seven putative sufficient conditions for the fairness of a selection procedure: (1) equality of test score means, (2) equality of the prediction systems, (3) the Einhorn-Bass Equal Risk Condition, (4) the Thorndike Constant Ratio Condition, (5) the Cole Conditional Probability Condition, (6) the equal probability condition, and (7) the Darlingotn Culture-Modified Criterion Strategy.

Educational researchers and psychometricians, however, have been unable to agree upon whether one of these conditions is the proper condition for the fairness of a selection procedure. In the light of this disagreement, the decision-theoretic approach has been put forward.

Some of the advantages (and disadvantages) of the approach have been discussed by Linn. As Linn has recently written, "the lack of general agreement on what constitutes fair test use creates real discomfort for anyone who would like to have a well-specified set of procedures to follow in order to insure fairness" (Linn, 1975b, p. 40). It should be clear that there are two logically distinct questions here. First, what is fairness? That is, what are the necessary and sufficient conditions for fairness? Second, how can one make a procedure which will lead towards the achievement of selection fairness? That is, how can one bring about the conditions for fairness? Linn went on to say that "this desire for a definite and mechanical answer may, in fact, be part of the problem, however. The question of what constitutes fairness or how admissions decisions should be made (in order to insure fairness) for members of different groups apparently is not amenable to a strictly technical solution" (p. 40). As one considers these issues, one finds that "in deciding how to pursue the goal of increasing the number of minority persons in law schools and the goal of admitting to the law school those students with the highest probability of excellent academic performance, it is necessary to look beyond a technical resolution of the problem. Competing values must be weighed" (p. 40). It is claimed that the decision-theoretical approach is useful in these situations because "the decision theoretic models provide a formal procedure for taking values into account" (Linn, 1975a, p. 194). According to Petersen and Novick, "the advantage of this model is that it requires an explicit public statement of utilities for each subpopulation. If I do not like the utilities you provide, an informative debate will no doubt ensue, with neither of us claiming any axiomatic justification of our utilities.

The important thing is that the discussion be public with all interested parties participating in the debate" (1976, p. 28). It is claimed, however, that the decision-theoretic approach has special importance. "The most satisfactory means to weigh the competing values explicitly is to formulate the admissions problem in decision-theoretic terms" (Linn, 1975b, p. 41 and Novick & Petersen, 1976, p. 86).

Petersen and Novick, allegedly building upon the work of J. Von Neuman and O. Morgenstern (1947) and A. Wald (1950) were among the first to bring the decision-theoretic approach to bear on the issue of test fairness. They suggested that there is a connection between rationality and decision theory. They wrote that:

> In the theory of the rational-economic man, developed in these writings, when all probabilities of outcomes are assumed known . . . , there is a simple paradigm required for rational decision. In that paradigm the desirability or utility of each possible outcome is stated quantitatively. Then, given all available information concerning the person in question, the probability of each possible outcome is stated for each decision under consideration. Next, for each possible decision, the utility of each outcome is multiplied by the probability of each outcome and the products are summed to provide an expected utility. Finally, that decision is then made for which the expected utility is highest. (1976, p. 23)

Petersen and Novick also made the following remarkable claim: "it would seem that there can be no distinction between maximizing expected utility and strategy fairness" (1974, p. 45). As I shall argue, this claim is completely unjustified. (See also, Novick & Petersen, 1976, p. 86.)

Some have pointed out that there are certain difficulties with the approach. For example, it has been warned, "Make no mistake. The issues will not be settled by mathematical specialists. Utility theory cannot crank out a just answer to a question of resource allocation,

when the parties with conflicting interests do not agree on the pay off matrix to be used or when the payoffs of the parties cannot be averaged" (Cronbach, 1976, p. 31). It must be recognized that "Procedures for involving different interest groups in the process, for translating value judgments into utilities, and for adjudicating differences in utilities assigned by different groups will be needed" (Linn, 1976, p. 56). Apparently, the value judgments are subjective, but once these judgments are known, the decision-theoretic approach can calculate the solution.

A similar point has been made by Novick and Ellis. They wrote that:

> Different decision-makers will likely have different utilities. Presumably, the utilities of the applicant will differ from those of the institution. . . . To implement the theory, it may be necessary to arrive at an agreement concerning utilities or to decide that one person's should determine the issue. However, it may not. Petersen (1976) has demonstrated that two widely different utility functions will lead to the same decision for many applicants. (1977, p. 319)

They went on to note that:

> When it is necessary to resolve differences in utilities, various techniques are available . . . it is necessary to study proffered utilities, to discover and remove discomforting implications (incoherencies) by modifying the stated utilities and to obtain some degree of consensus among those concerned. (1977, p. 319)

It is also recognized that there are some difficulties with decision-theory's assumption of "additivity." As Linn has noted, "In the standard approach used in decision-theory it is assumed that the utility of an outcome for one individual does not depend on what happens to other people. This is known as the assumption of additivity and is used in the threshold utility model by Gross and Su (1975). It may sometimes be more realistic, however, to allow the utility for one individual to

depend on who else is selected (Linn, 1975a, pp. 199-200). L. J. Cronbach has written that "alternative axiom sets can embody views so divergent as to defy comparison. Although weak, nonadditive axioms are much used in economic decision theory, every paper in the psychometric literature relies on the strong additive axiom" (1976, p. 41).

There are prima facie difficulties with this approach. First, it needs to be determined whose utilities are to be taken into account. Surely the utilities of the institution and the utilities of the applicants from each group should be considered. It also seems reasonable that the utilities of other people should be considered, especially those people who may be indirectly affected by the policies of the institution. It seems that morality would require either equal consideration of interests or equal consideration of those whose interests were substantially affected.

At this point some things should be said about the psychometrician's desire to reach a consensus on the utilities of the possible outcomes. I agree that we should always try to reach a consensus, but the fact of the matter is that many important situations are characterized by conflicts of interest. Indeed, Baier has said that morality is the court of appeal for conflicts of interest (1965, pp. xvi, xix). Rawls has written that justice is essentially concerned with the elimination of arbitrary distinctions and with the establishment, within the structure of a practice, of a proper balance between competing claims (in Bedau, 1971, p. 77). He has written that, "justice is the virtue of practices where there are assumed to be competing interests and conflicting claims and where it is supposed that persons will press their rights on each other" (Rawls, 1971, p. 134). The principles of justice are principles

which justify ways of adjudicating such competing interests and conflicting claims. It is typical of cases of justice to involve persons who are pressing on one another their claims, between which a fair balance or equilibrium must be found (in Bedau, 1971, p. 83).

This brings me to my third point. The question that is before us is whether a certain selection procedure is fair. The decision theorist is committed to take a position very similar to that which a "naive" utilitarian would take. He is likely to use the total utility (or the average utility) as a measure of the policy's desirability. The total utility is an additive function of the utilities of all individuals, where each person's utilities are given identical weight. Now, it may happen that the total utility is maximized when many people are well off and a few are quite disadvantaged. The principle of total utility would allow the disadvantages of a few people to be outweighed by the greater advantages of others. Whatever the moral status of the principle of total utility, the principle of justice (or fairness) is distinct from the principle of utility. The principle of justice would not allow the disadvantages of some to be outweighed by the greater advantages of others (Rawls, 1971, pp. 80-99). Justice has to do, not so much with the quantity of good or evil, as with the manner in which it is distributed (Frankena, 1962, p. 4; Frankena, 1973, p. 43; and Barry, 1965, pp. 43-44). Thus, it is pointed out that justice is not an additive (or an aggregative) principle, but that it is a distributive principle. This is a truth about the logical properties of the concept of justice.

The decision-theorist is susceptible to a similar kind of criticism. We are not seeking a policy which maximizes the expected utility but a policy which is fair. It can turn out that a policy which maximizes the

expected utility is not fair. Even if one could establish that a policy maximized the expected utility of the relevant parties, such a fact would not be sufficient to show that the policy treated people fairly. One needs to show that the policy brought about a certain kind of distribution.

Now the brief comments given above are not meant to resolve these disputes, but only to locate more precisely what these disputes are about. I shall present some general criticisms of decision theory in the section on General Remarks.

Specific Applications

(a) The threshold utility strategy. There may be situations in which an institution is legitimately interested in distinguishing between those applicants whose performance will be successful and those applicants whose performance will not be successful. That is, there might be a minimum level of criterion performance so that the institution would find its purposes (best) served if all of those whose qualifications fell above the minimum were treated the same, and all of those whose qualifications fell below the minimum were treated the same. Ideally, the institution would like to accept those who are above the minimum and reject those below the minimum. The institution may also find that it can assign values (or utility or desirability) for each group (i.e., the majority and the minority) to each of the four selection outcomes: accept the qualified, accept the not qualified, reject the qualified, and reject the not qualified. Given the probabilities based on the criterion-test relationship, the selection ratio, the number of applicants, and the number of positions, the institution tries to find the test score cut point for each group which will maximize the expected

utility. The test score cut point divides the applicants into those who will be accepted and those who will be rejected (on the basis of test scores). Apparently, the selection procedure is fair if the expected utility of the final decision is the same for each group. (Petersen & Novick, 1974, p. 41; Linn, 1975a, p. 196; and Linn, 1975b, p. 41).

(b) The expected utility strategy. Having "determined" the probabilities and utilities, the institution computes the average or expected utility for each person and then selects that person whose performance has the higher expected utility. Apparently, then, the selection procedure is fair if people who have the same expected utility are treated the same (that is, either rejected or accepted). Consider the following example for applicants A and B. Notice that B's utilities are higher than A's utilities. The test user (i.e., the institution) has placed extra value on B's being disadvantaged, say (Petersen & Novick, 1977, pp. 318-319; and also Cronbach, 1976) One calculates that A's expected utility is .530 while B's expected utility is .545 and therefore one concludes that B should be chosen instead of A.

Insert Figure 4 about here

Specific Criticism

(a) The threshold utility strategy. Remember, the institution is interested in distinguishing between those who are, say, suitably qualified and those who are not suitably qualified. The test score cutting point for each group draws the boundary line between the test scores for applicants who are accepted and the test scores for applicants who are rejected. Also, values (or utility or desirability) are (subjectively)

attached for each group to each of the four outcomes (accept a suitably qualified, accept a not suitably qualified, reject a suitably qualified, reject a not suitably qualified). Given the selection ratio, the number of applicants, and the number of positions, then the expected utility is maximized by adjusting the test score cut-points for each group. The relevant probabilities are based on the criterion-test relationship in terms of predictive validity, so perhaps the probabilities need not be subjective.

Suppose that the appropriate utilities have been taken into account. Suppose that the cutting points have been found which maximize the expected utility, and that the expected utility is the same for each group. Is the selection procedure fair? No; both the principle of equal opportunity for those suitably qualified and the principle of equal consideration are violated. Since the arguments are identical to those used, for example, in the criticism of the regression coefficients, I shall omit repeating them. (Remember, the Threshold Utility Strategy is built upon the criterion-predictor relationship.)

There is an additional problem with this application of decision-theory to selection procedures. The threshold-utility strategy really accomplishes only intrapolicy optimization. The optimization presupposes that the criterion-test relationship is to form a legitimate part of the selection procedure. But, whether the criterion-test relationship should be used is also open to determination by decision theory. For example, the consequences of a policy which used the criterion-test relationship should be compared with the consequences of alternative policies and, in particular, with the consequences of the open admissions policy to see which has the greater expected utility. And although the

open admissions policy would have lower efficiency, it might well be (more) fair and less expensive to administer. If institutions of higher education had open admissions policies, then high school students might view their going to college in a different light. For one thing, fierce competition would not help one get into college. Of course, a student would be well advised to be prepared for the possible opportunity. Also, the widespread use of open admissions might provide scientifically useful data about colleges and students, since the students would be selected by lot. Now, I don't want to belabor the point. But, on the principles of decision-theory, not only should intrapolicy procedures be evaluated, but all alternative policies (and strategies) should be evaluated. The policy which is "justified" by the principles of decision-theory is the one which maximizes the expected utility. It may well be that the open admissions procedure would lose out to the current policies that use the criterion-test relationship, but this has to be shown by collecting the empirical evidence and calculating the relevant expected utilities. It is an important factual premise which cannot merely be conceded without argument. Also, it would be interesting to see how open admissions compare to the current policies. If the expected utilities were reasonably close, the fact that open admission is (more) fair may decide the question in favor of open admissions.

The expected utility strategy has the same assumptions and presuppositions as the threshold-utility strategy, except that it would not divide the applicants into a group whose scores were above the cutting point and into a group whose scores were below it. The expected utility strategy would evaluate each applicant with respect to his probability of achieving various criterion scores given that he has a certain test

score. Again, it can be shown that the expected utility strategy is subject to the same difficulties as the threshold utility model; it violates the principles of fairness. Thus, the expected utility strategy provides us with neither a necessary nor a sufficient condition for fairness.

General remarks. Turning from offering criticism of the specific recommendations based on decision theory, I would now like to offer some general remarks about the application of decision-theory to questions of rationality and justice. An adequate discussion of decision theory would clearly require much more space than the amount I shall devote to it in this essay. My discussion will provide only a glimpse of some of the important issues concerning the application of decision theory. I shall point out, however, certain prima facie problems that a more adequate discussion should take into account before decision theory can be applied in the name of rationality or justice.

Rationality, Fairness, and Maximizing One's Expected Utility

It is helpful to make some distinctions among possible uses of the term "rationality." I shall refer to these as the technical, the descriptive, and the normative uses.

First, let me explain the technical use. It may be that economists can enrich their economic theory by stipulating that rational behavior is that behavior in which the agent maximizes his expected utility. For example, when Lange says that "a unit of economic decision is said to act rationally when its objective is the maximization of a magnitude," he is stipulating the definition of the technical term "rational behavior" (1945, p. 30). Such definitions may enable economists to discover new and broader generalizations about the economic behavior of people in

society. Also, economists may find it useful to adopt general postulates. For example, "the postulate of rationality is the assumption that all units of economic decision act rationally" (p. 30). Lange writes that this assumption is "a most powerful tool for simplification of theoretical analysis" and that it is "a short-cut to the discovery of laws governing the decisions of units and to the prediction of their actions under given circumstances" (p. 30). In such situations, one's use of the term "rationality" is justified insofar as the term is helpful in developing and advancing economics as an empirical science. That the actual behavior of people who are described as rational in the ordinary sense may not be adequately described as cases of maximizing one's expected utility is irrelevant when one is considering the term's status in certain kinds of scientific inquiry. The technical term is introduced in a theory for purposes of explanation and prediction. That the technical term "rationality" is logically distinct from the ordinary use of "rationality" should not be surprising when one realizes what the theoretician is trying to do.

Some writers, however, seem to believe that it is an empirical generalization that rational behavior is behavior in which a person maximizes a magnitude (Lange, 1945, p. 30). The remarks by Petersen and Novick, which I have quoted earlier, imply that such an empirical relationship holds between rationality and maximizing one's expected utility. Thus, they appear to be using the term "rationality" in a descriptive way. Of course, before one can determine whether such claims are justified, one needs an adequate, logical analysis of the ordinary concept of rationality so that the scientist can go out and see if the empirical relationship holds. Useful steps in the logical analysis of rationality

have been taken by Black (1974), Mortimore (1976, Chs. 4, 15), and (Richards, 1971).[4]

The empirical evidence seems to support the conclusion that the technical models of rational choice do not explain or predict the actual behavior of people. Guttentag concluded that the hypothetical man presumed by decision-theory is not supported by empirical psychological data (1974, p. 596). Benn and Mortimore, after an extensive review of the technical models and their role in theorizing, concluded that the practical payoff in explanation and prediction of such theoretical ingenuity is frequently disappointing (1976, p. 190).

There are other difficulties with the view that rational behavior is behavior in which one maximizes his expected utility. Jeffrey put it quite nicely when he said:

> To deliberate is to evaluate available lines of action in terms of their consequences, which may depend on circumstances that the agent can neither predict nor control. . . . In the Bayesian model the agent's notions of the probabilities of the relevant circumstances and the desirabilities of the possible consequences are represented by sets of numbers combined to compute an expected desirability for each of the acts under consideration. The Bayesian principle for deliberation is then to perform an act which has maximum expected desirability. . . . The Bayesian model may be as applicable to the deliberations of a knave or a fool as to the deliberations of a good and wise man, for the numerical probabilities and desirabilities are taken to be subjective in that they reflect the agent's actual beliefs and preferences, irrespective of factual or moral justification. (1965, p. 1)

Another noted theorist, Harsanyi, has said:

> Utility theory hardly tells us more than the implications of the assumption that people maximize whatever they maximize, and game theory tells us what happens if several people acting consistently are at the same time trying to maximize whatever they are trying to maximize. (1966, p. 18)

Benn and Mortimore make a similar point when they argue that the technical models of rational choice give no importance to whether the

agent's postulated substantive ends are in any sense rational or not (1976, p. 190).

The thrust of these remarks is that decision theory is not a theory of rationality at all, but rather that it is a theory of deliberation. Such a theory apparently makes no distinction between the deliberations of a fool and the deliberations of a good or wise man. That is, decision theory gives a rather formal or abstract theory of deliberation. The theory is remote from the realities of the world. Clearly, additional principles are needed to specify and distinguish between irrational deliberation and rational deliberation. Maximizing one's expected utility is compatible with acting quite irrationally.

Some people have held, as does Raiffa, that decision theory does not present a descriptive (or empirical) theory of actual behavior, but that the theory prescribes how to act in a reasonable and practical way (i.e., a rational way). Raiffa has written that:

> The so-called Bayesian viewpoint uses (subjective) utilities and subjective probabilities . . . but the theory is not a description theory of actual behavior. . . . Rather the approach prescribes how an individual who is faced with a problem of choice under uncertainty should go about choosing a course of action that is consistent with his personal basic judgments and preferences. He must consciously police the consistency of his subjective inputs and calculate their implication for action. (1968, pp. x-xi)

According to this view, then, people are not behaving in a way which maximizes their expected utility, but people ought to do so.

Why should they do so? Raiffa has said that it would be a reasonable and a practical way of acting. At this point one should recall the remarks of Jeffrey, Harsanyi, Benn, and Mortimore--decision theory cannot distinguish between the deliberations of a knave or a fool and the deliberations of a good and wise man. The utilities (or desirabilities)

are taken to be subjective in the sense that they reflect the agent's actual beliefs and preferences, irrespective of factual or moral justification. A theory which cannot make such distinctions surely has but limited value in providing practical guidance for the way we should act. If the theory is to be used to provide practical guidance, the theory will need to be supplemented with principles which bring it into touch with the realities of decision making in the world.

Others have gone farther by arguing that the decision-theoretic approach cannot provide an adequate model for rational behavior. Richards has argued that in the process of deliberation, it is normally impossible, given the costs of prolonged decision making, to consider all the possible alternative plans and to choose the maximal one in light of all the relevant facts (1971, p. 46). Simon has suggested that a rational person is not required to find the maximal alternative but only one which is satisfactory. An alternative is satisfactory if (1) there exists a set of criteria that describes the minimally satisfactory alternatives, and (2) the alternative in question meets or exceeds all the criteria (1957, p. 204). Black has argued that judgments about an action's reasonableness are typically comparative; if we can sometimes say that one action is more reasonable than another, we are very seldom able to choose a single action as uniquely reasonable (1972, p. 52). Such a view is relevant to questions about the nature of rationality because reasonableness is closely related to rationality.

It may be suggested that since decision theory is normative, it can provide us with moral guidance. Remember, however, that decision theory cannot distinguish between the deliberations of a knave or a fool and the deliberations of a good and wise man. An agent's notions of the

desirabilities of the possible consequences are subjective in the sense that they reflect the agent's actual beliefs and preferences, irrespective of factual or moral justification. In other words, in such a conception, decision theory is morally neutral. Being guided by decision-theory is compatible with following policies and procedures which have vicious consequences. On the other hand, to say that performing an action is fair is to imply that, so far, the action is justified. To say that an action is fair is to imply that there are good reasons, though perhaps not conclusive reasons, for performing the action.

Let us suppose that, contrary to fact, decision theory has provided us with a normative theory of rationality. Still, the theory would not provide an adequate basis for a theory of morality. Rationality is logically distinct from morality. Morality requires something more than or other than rationality in that it delimits certain ends (Baier, 1965, Ch. 5.3; Richards, 1971, p. 76). It can also be shown that moral reasons have supremacy or priority over rational reasons (Baier, 1965, Ch. 7.3; Richards, 1971, Ch. 12).[5]

Finally, justice is only one species of morally right-making considerations (Frankena, 1962, p. 5). In particular, the principle of social utility is logically distinct from the principle of justice. The principle of utility is an aggregative principle, but the principle of justice is a distributive principle. Justice has to do, not so much with the quantity of good or evil, as with the manner in which it is distributed (Frankena, 1962, p. 4).

Petersen and Novick are quite mistaken, therefore, when they claim that there can be no distinction between maximizing expected utility and strategy fairness (1974, p. 45). They are unjustified in claiming that

a formal decision-theoretic approach should be used in reaching educational decisions (Novick & Petersen, 1976, p. 86). The truth of the matter seems to be that (formal) decision theory cannot distinguish between rational and irrational decisions nor can it distinguish between fair and unfair decisions. If one wants to be fair, he must look for guidance beyond decision theory. In order for the theory to be useful in providing moral guidance, the theory needs to be supplemented with principles which make the distinctions between right and wrong and between fair and unfair.

Possible modifications. I propose now to show two ways in which the decision-theoretic position might be modified in order to become a guide for moral decision making.

Let us consider the following case. A black applicant and a white applicant are equally qualified for entrance to medical school. It seems that fairness would require that equal treatment should be given to those with equal qualifications. It happens to be the case, however, that the social value of a black student becoming a doctor is greater than the social value of a white student becoming a doctor. Perhaps there is a high demand for black doctors, but there is a small supply. The college admissions officer might decide, after considering these facts, to choose the black applicant over the white applicant. The officer might defend his decision by pointing out that a greater social value is being pursued.

There is a difficulty here, however, for the officer has ignored certain probabilities in reaching his decision. In particular, suppose that it is the case that 90% of the black students flunk out of medical school, while 90% of the white students eventually succeed in medical

school. It may well turn out, then, that selecting the white applicant eventually brings about larger expected utility than selecting the black applicant, and, therefore, that the admissions officer had been unfair to the white student, here, too. An extreme case would be the one where all of the black students fail, while all of the white students succeed. Should we say that the admissions officer has merely made an error in calculation? No. It seems that he has made a moral mistake, a mistake which is particularly offensive in the extreme case presented above. The admissions officer has violated the moral principle which requires that an agent attend to the maturation of future utility in reaching his decisions.

Again, suppose that the black and the white applicant are equally qualified. Let us suppose that the white student's chances of succeeding in medical school are 60%, while the black student's chances are 50%. One is able to make these judgments of probability because people have collected reliable data about these kinds of students. The college admissions officer might decide, after considering these facts, to choose the white applicant over the black applicant. The officer might defend his decision by pointing out that the white applicant is more likely than the black applicant to succeed in college.

There is a difficulty, however, for the officer has ignored certain utilities. Supposing the disutilities are the same, because the social need for black doctors is so much greater than the social need for white doctors, the expected utility for the black applicant is much larger than the expected utility for the white applicant. An extreme case would be the one where there are no jobs at all for white doctors. Should we say that the officer has made a mistake in calculation? No.

It seems that he has made a moral mistake, that he may be guilty of moral myopia. For he does not care about what happens to the applicants or to society after the students finish college. Morality requires, however, that a person care about the consequences of his actions on others. A moral agent must attend to the total consequences of his actions in reaching his decisions. Thus, one might say that a moral decision-maker must be guided by the principle of total consequences.

Remember, decision theory holds that the agent's notions of the probabilities of the relevant circumstances and his notions of the desirabilities of the possible consequences are subjective in the sense that they reflect the agent's actual beliefs and preferences, irrespective of factual or moral justification. The two examples given above, however, show that a moral decision maker must be guided by the principle of the maturation of future utilities and the principle of total consequences. If a decision maker fails to attend to such principles, then he is, so far, behaving immorally.

More can be said about the principles to which a moral agent must attend, but I shall leave these matters to the side. (For further discussion, see Baier, 1967; Hare, 1963; Rawls, 1971; Richards, 1971.)

A more fair policy.[6] I turn now to presenting an argument for a policy which is morally better, in particular, more fair, than the policy which decision theory would dictate.

Let us consider the following case. Persons A and B are in equal need of medicine, but only one unit of the drug is available. Suppose that A's chances of being cured by the drug were 100% while B's chances of being cured are 0%. It seems it would be fair to give the drug straight-away to A and not to B. After all, if A were not around, it

would be a waste of the medicine to give it to B. If A were not around, it would be senseless to waste the drug by giving it to B.

Now, suppose that A's chances of being cured by the drug are 83%, while B's chances of being cured are 80%. One might decide to give the drug to people with the higher chances of being cured because, in the long run, such a policy would be likely to cure more people. It must be admitted that this is a good reason for pursuing such a policy. On the other hand, there are good reasons for treating A and B the same. And since there is only one unit of the drug, it would be fair to decide who is to receive it by drawing lots. First, there is the practical difficulty that there are errors of measurement. Our estimates of A's and B's chances of being cured by the drug are too close to each other. It would be wrong to pay attention to such small differences. The difference between A's and B's estimates are too small for practical decisions. Therefore, A and B should be regarded as being the same. Secondly, suppose that A's chances <u>really</u> are 83% while B's chances really are 80%. Even here it seems it would be fair to treat them as equals. After all, if A were not around, then it would be a good use of the medicine to give it to B. It would surely not be a waste of the medicine to give it to B.

It seems that if a person's chances of being cured were below some minimum level, it would be proper to regard giving him the medicine as a waste of the medicine. On the other hand, it seems that if a person's chance of being cured were above some higher level, then it would be proper to regard giving him the medicine as a good use of the medicine. If there were several people above the higher level, then it would be fair to treat them as equals. After all, the medicine would be put to

good use if it were given to any of those whose chances were above the higher level.

There are many people whose chances of being cured are greater than the minimum level but less than the higher level. For such people their chances are not so low that it would be a waste to give them the medicine, yet they may not be so high that it would be a good use of the medicine to give it to them. After taking the errors of information into account, it seems that this middle region should be divided into three regions: (1) The first would contain those whose chances are within the measurement error of being equal to the higher level. Such people should be treated as equal to those whose chances are above the higher level. (2) The second would contain those whose chances are within the measurement error of being equal to the minimum level. Such people should be treated as equal to those whose chances are below the minimum level. (3) The third would contain all the rest of the people whose chances fall somewhere in the remaining, middle range.

It seems to me that the following policy would be morally better and, in particular, more fair than the policy dictated by decision theory. A decision theory policy would distribute the medicine to those in the middle group only when all those in the high group have received the medicine. In the face of scarcity, people in the middle group would probably not receive medicine. My recommendation is this. Given that there is a certain amount of medicine, one should give the medicine to those in the high group, except that one should distribute 25%, say, of the medicine in a fair way to those in the third part or the middle group. This policy is morally preferable to the decision-theoretic policy. Such a policy would be reasonably efficient in putting the

medicine to good use, and it would give those people in the third part or the middle group some chance of obtaining the medicine which has a moderate chance of curing them.

I believe a similar recommendation is appropriate for college admissions procedures. Such a policy would be more fair than most of the current policies and more fair than the decision-theoretic policy. First, there are many more applicants than there are openings; there is a scarcity of openings. Second, the empirical predictors of an applicant's success are but roughly accurate. Nonetheless, the empirical predictors can be used to create a policy that is more fair than current procedures. The predictors can be used to tell us the applicant's chances of success.

It seems that if the applicant's chances of success were below some minimum level, then it would be proper to regard giving him the opportunity at a college education as a waste of the opportunity. On the other hand, if the applicant's chances of success are above some higher level, then it would be proper to regard giving him the opportunity as a good use of the institution's resources. If there were several people above the higher level, then it would be fair to treat them as equals. After all, an opportunity would be put to good use if it were given to any of those whose chances were above the higher level.

There are many people whose chances of success are greater than the minimum level but less than the higher level. For such people, their chances are not so low that it would be a waste to give them the opportunity, nor yet so high that it would be a good use of the opportunity to give it to them. After taking into account the errors of information and making the three-part classification as in the medicine example, the policy will give attention to those in the (remaining) middle region.

Given that there is a scarcity of opportunities, one should give the opportunities to those in the high group, except that one should distribute 25%, say, of the opportunities in a fair way to those in the middle group.

Such a policy is morally preferable to the decision-theoretic policy. The policy would be reasonably efficient in putting the institution's resources to good use, and it would give those in the middle group some chance of obtaining an opportunity which they have a moderate chance of using profitably.

Clearly, the determination of the minimum level, the higher level, and the percentage of opportunities to be set aside for the middle group will have to rely upon intuitive and informal judgments. It is hoped that the people can agree that certain judgments about such things are reasonable.

Conclusion. As I have noted above, it has been urged, on behalf of the decision-theory approach, that it is the "most satisfactory means to weigh the competing values explicitly." I have suggested that attempts to specify the relationship between rationality and decision-theory in an empirical or a normative sense is beset with prima facie difficulties. Also, I have argued that the concept of fairness is logically distinct from the concept of maximizing expected utility, and I have argued that two specific recommendations based on decision-theory which use the criterion-test relationship would be, as a matter of fact, fair in only the grossest sense. Furthermore, it seems that our understanding of these issues would be greatly improved by trying to find good reasons for pursuing certain institutional purposes and by trying to delimit the important intended (and unintended) consequences of the various policies.

It would be helpful to understand why, and to what degree, it is important to increase the representation of certain minority groups in the institution. There is still a lot of room for argumentation in these matters, for we have not reached the level of fundamental principles of moral reasoning. It seems that these activities would go a long way towards increasing our understanding of the issues and also would lead us in the direction of justifying certain policies. Here I am applying the method of moral debate.

Decision-theory may, indeed, have a contribution to make to our understanding of these issues. But it is not the only means available, and it may not be the most satisfactory. A supporter of decision theory needs to show in specific ways the kind of contribution the decision-theory approach can make to increasing our understanding of these issues. That dangers are present in applying decision-theory should be clear by now.

In conclusion, I have suggested that the claims that the decision-theoretic approach may help us to determine a fair selection procedure are unfounded. The concept of fairness is logically independent of the concept of maximizing one's expected utility as it is conceived by decision-theorists. To pursue certain decision-theory recommendations which use the criterion-test relationship would be, as a matter of fact, fair in only the grossest sense.

Reflection

Suppose, as many educational researchers and psychometricians have come to believe, that the prediction equations which use standardized tests and grade point averages are the best that it is empirically possible to obtain. (Remember, the size of the correlation coefficients

is about .58 in undergraduate selection, .40 in graduate selection, and .30 in employment selection.) In the philosophical literature, writers have worried about whether social justice (or fairness) might be overruled by social utility (or by the greatest happiness of the greatest number). In other words, the writers have worried that social justice may not be preferred in relation to social utility. What the facts about current selection procedures indicate is not that social utility might be chosen over fairness, but that fairness is not empirically achievable when one uses current prediction techniques, although such techniques turn out to be quite efficient.

Perhaps it could be shown that these prediction techniques are better in terms of fairness than any other selection procedure. If that were the case, then one s using the test would be morally regrettable, in a sense, but not morally censurable for it does help secure some degree of fairness. Perhaps it could be shown that these techniques are, on balance, the morally right techniques to use. Whatever the logical status of these arguments, the selection procedure would still be fair in only the grossest sense. And it should be recognized as being grossly unfair, for only then will a better alternative be sought and perhaps found. One should not call our "best effort" fair; that only distorts the concept of fairness and misleads people about the important moral characteristics of the selection system.

These remarks are similar to remarks about social justice which were made by Gregory Vlastos (1962, p. 41). He suggested that only an "affluent" society could afford social justice. That is, only in special situations is it possible to achieve social justice. Thus, if the conditions are not met, the society cannot be, and is not, just, although

it may not be liable to moral censure. Of course, a society should be trying to achieve social justice. It would be a mistake, perhaps a pernicious one, to call a society just when, in fact, it has not achieved nor made serious efforts to achieve social justice.

If my criticism is sound, then serious attention should be given to determining whether there are ways in which fairness might be achieved. Our society holds fairness to have considerable moral value. Perhaps, in the light of such gross unfairness which results from using current prediction systems, the value of fairness may override the value of efficient student selection so that institutions of higher education should move toward open admissions. After reconsidering the purposes which our institutions are pursuing, the value of education to the student and to society, and the facts about the fairness and the efficiency of the alternatives, the position of the radical egalitarian (i.e., that there ought to be open admissions) may not seem so radical. Perhaps my recommendation will, after careful reflection and debate, be chosen because it is reasonably efficient and yet more fair than the alternatives. I am inclined to prefer such a policy over the radical egalitarian's open admissions policy, but further discussion is needed.

One might feel that a society has no use for principles which cannot be presently achieved in social policy. This would be a mistake. Certain principles of morality (and rationality) serve as <u>standards of achievement</u> against which the actual behavior of social policies may be measured and judged. They serve as <u>ideal directives</u> to which humankind should strive to make its institutions conform. Analogously, it is a misunderstanding to claim that the principles of rational and logical thought have no value because many people fail to conform to them.

Conclusion

In this essay I have addressed the aspect of the selection procedure process in which unfairness may be brought about by the role of certain practical predictors or indicator standards in selecting applicants. Whether these predictors are useful depends in part upon their empirical relationship with the criterion measures (which, in turn, are related to the theoretical qualifications).

I have assumed that the institutions have good reasons for pursuing their (global) purposes (Assumption I). Also, I have assumed that the institutions have objectively determined the qualifications which students must have in order for these institutions to achieve their global purposes, and that these (theoretical) qualifications have a reliable, unbiased, perfectly relevant criterion measure which can be operationally defined as the first-year grade point averages (Assumption II). Although these assumptions need justification, I wanted to point out that the selection procedure is liable to criticism at another aspect; that apsect which is concerned with the role of the empirical predictors of criterion performance.

I began by showing the distinction between procedural fairness and substantive fairness. Later, I suggested that there are three substantive principles of fairness which are applicable to college selection procedures: the principle of equal chance for all, the principle of equal opportunity for those suitably qualified, and the derivative principle of equal accuracy.

I then considered eight general positions which have claimed to provide a sufficient condition for the fairness of a test in its role in the selection procedure. These are: (1) equal test score means for the

relevant groups, (2) equality of prediction systems, (3) the equal risk condition, (4) the Thorndike constant ratio condition, (5) the Cole conditional probability condition, (6) the equal probability condition, (7) the culture-modified criterion approach, and (8) the decision-theoretic approach. I have argued that none of the positions states a sufficient condition, that the Thorndike condition is a necessary condition according to the principle of equal chance for those suitably qualifed in special situations, and that the Cole condition is a necessary condition according to the principle of equal accuracy in special situations.

Using Assumptions I and II, I argued that educational researchers and psychometricians have not captured the essential conditions for a test to be fair in its role in the selection procedure.

I have also argued that current selection policies which are based upon the "best available" prediction equations are remarkably efficient but that they are fair in only the grossest sense according to all three of the substantive principles of fairness. Whether or not following current practices can be shown to be morally right, it is important to recognize that these policies are unfair to both minority and majority applicants.

Finally, I have recommended a policy which uses the "best available" predictors, but which, as I have argued, is reasonably efficient and more fair than the current policies.

One final comment. In this essay I have been investigating whether certain conditions are necessary or sufficient for the fairness of the test (or empirical predictor) in its role in the selection procedure. I have been investigating whether a particular kind of unfairness may correctly be attributed to the test in its role in the selection procedure.

One must remember that there are other kinds of unfairness that a selection procedure can bring about. It is most important that we determine whether the institution's purposes and its related qualifications are themselves just. And here we must be concerned with questions of compensatory justice and with questions about the institution's role in the distribution of social and economic benefits and burdens. Such questions belong to the realm of political and moral philosophy and, at the present, the only method we have for dealing with such questions is moral deliberation and debate. There is no moral calculus, no moral geometry; moral debate relies on moral reasoning and judgment.

1FSE:a

Bibliography

Albemarle Paper Co. v. Moody. 422 U.S. 405 (1975).

Anastasi, A. Psychological testing (3rd. ed.). New York: Macmillan, 1968.

Archibald, K. (Ed.). Strategic interaction and conflict. Berkeley: University of California, 1966.

American College Testing Assessment Program. Assessing students on the way to college (Technical Report, Vol. 1). Iowa City, Iowa: ACT Publications, 1973.

Bakke v. The Regents of the University of California. San Francisco, 23311.

Bakke v. The Regents of the University of California. San Francisco, 23311, Dissenting Opinion by J. Tobriner.

Baier, K. The moral point of view. New York: Random House, 1965.

Barry, B. Political argument. New York: Humanities Press, 1965.

Bass, A. R. The "equal risk" model: A comment on McNemar. American Psychologist, 1976, 31, 611-612.

Bass, A. R., & Einhorn, N. J. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.

Bedau, H. A. (Ed.). Justice and equality. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Bedau, H. A. Radical egalitarianism. In H. A. Bedau (Ed.), Justice and equality (pp. 163-180). Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Benn, S. I., & Mortimore, G. W. (Eds.). Rationality and the social sciences: Contributions to the philosophy and methodology of the social sciences. London: Routledge & Kegan Paul, 1976.

Benn, S. I., & Mortimore, G. W. Technical models of rational choice. In S. I. Benn & G. W. Mortimore (Eds.), Rationality and the social sciences (pp. 157-195). London: Routledge & Kegan Paul, 1976.

Berlin, I. Equality as an ideal. In F. A. Olafson (Ed.), Justice and Social Policy (pp. 128-150). Englewood Cliffs, N.J.: Prentice-Hall, 1961.

Black, M. Rationality and cultural relativism. In M. Black (Ed.), Problems of choice and decision: Proceedings of a colloquium held in Aspen, Colorado, June 24-July 6, 1974 (pp. 128-160). Ithaca, N.Y.: Cornell University Program on Science, Technology, and Society, 1975.

Black, M. Reasonableness. In R. F. Dearden, P. H. Hirst, & R. S. Peters (Eds.), Reason: Part 2 of education and the development of reason (pp. 44-57). London: Routledge & Kegan Paul, 1972.

Blackstone, W. T. Reverse discrimination and compensatory justice. Social Theory and Practice, Spring 1975, 3(3), 253-288.

Boxill, B. R. The morality of preferential hiring. Philosophy and Public Affairs, 1978, 7(3), 246-248.

Brandt, R. B. (Ed.). Social justice. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Breland, H. M. Defunis revisted: A psychometric view. Princeton, N.J.: Educational Testing Service, August 1974.

Breland, H. M., & Ironson, G. H. Defunis reconsidered: A comparative analysis of alternative admissions strategies. Journal of Educational Measurement, Spring 1976, 23(1), 89-99.

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 4, 115-124.

Cole, N. S. Bias in selection. Journal of Educational Measurement, 1973, 10, 237-255.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, I., Mood, A. M., Weinfeld, F. D., & York, R. L. Equality of Educational Opportunity. U.S. Department of Health, Education & Welfare. Washington, D.C.: U.S. Government Printing Office, 1966.

Cronbach, L. J. Equity in selection--Where psychometrics and political philosophy melt. Journal of Educational Measurement, Spring 1976, 13(1), 31-42.

Darlington, R. B. Another look at "culture fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Darlington, R. B. A defense of "rational" personnel selection, and two new methods. Journal of Educational Measurement, Spring 1976, 8(1), 43-52.

Ellett, F. S., Jr. A more fair policy and the lionits of fair selection. Proceedings of the Annual Meeting of the Philosophy of Education Society, Toronto, April 1979.

Ellett, F. S., Jr. Decision theory, rationality and moral justice. Los Angeles: Center for the Study of Evaluation, University of California, Los Angeles, November 1979.

Feinberg, J. Social philosophy. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Fiss, O. M. Groups and the equal protection clause. Philosophy and Public Affairs, Winter 1976, 5(2), 107-177.

Flaugher, R. L. The new definitions of test fairness in selection: Developments and implications. Educational Researcher, October 1974, 3, 13-16.

Frankena, W. K. The concept of social justice. In R. B. Brandt (Ed.), Social justice (pp. 1-29). Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Frankena, W. K. Ethics (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Fullinwider, R. K. Preferential hiring and compensation. Social Theory and Practice, Spring 1975, 3(3), 307-320.

Goldman, A. H. Limits to the justification of reverse discrimination. Social Theory and Practice, Spring 1975, 3, 289-306.

Goldman, A. H. Affirmative action. Philosophy and Public Affairs, Winter 1976, 5(2), 178-195.

Goldman, A. H. Justice and hiring by competence. American Philosophical Quarterly, January 1977, 14(1), 17-28.

Griggs v. Duke Power Co., 401 U.S. 424, 1971.

Gross, A. L., & Su, S. Defining a "fair" or "unbiased" selection model: A question of utilities. Journal of Applied Psychology, 1975, 60, 345-351.

Gross, B. (Ed.). Reverse discrimination. New York: Prometheus Books, 1977.

Guion, R. U. Employment tests and discriminatory hiring. Industrial Relations, 1966, 5, 20-37.

Guttentag, M. Choice and decision: A review of relevant empirical work. In M. Black (Ed.), Problems of choice and decision, 1974.

Hare, R. M. Freedom and reason. Oxford: Oxford University Press, 1963.

Harsanyi, J. C. Some social-science implications of a new approach to game theory. In K. Archibald (Ed.), Strategic interaction and conflict. Berkeley: University of California, 1966.

Harvard College, Brief of the President and Fellows. Amicus Curiae in support of the respondents in the case of Defunis v. Odegaard, 416 U.S.

Heslep, R. D. Preferential treatment and compensatory justice. Educational Theory, Spring 1976, 26(2), 147-153.

Holman, M. G., & Doctor, R. Educational and psychological testing. New York: Russell Sage Foundation, 1972.

Hunter, J. E., & Schmidt, F. L. Critical analysis of the statistical and ethical implications of various definitions of "test bias." Psychological bulletin, 1976, 83(6), 1053-1071.

Jackson, G. D. On the report of the Ad Hoc Committee on Educational Uses of Tests with Disadvantaged Students: Another view from the Association of Black Psychologists. american Psychologist, 1975, 30, 88-93.

Jagger, A. Fairness, meritocracy, and reverse discrimination. Social Theory and Practice, 1977, 4(2), 211-235.

Jeffrey, R. C. The logic of decision. New York: McGraw-Hill, 1963.

Lange, O. The scope and method of economics. The Review of Economic Studies, 1945, 13, 19-32.

Lawyer's Committee for Civil Rights Under Law, Brief of Amicus Curiae in support of the respondents in the case of Defunis v. Odegaard, 416 U.S.

Linn, R. L. Fair test use in employee selection: The models and the evidence. Report prepared under the contract with the American Telephone and Telegraph Company, 1975(a).

Linn, R. L. In search of fair selection procedures. Journal of Educational Measurement, Spring 1976, 13(1), 53-58.

Linn, R. L. Test bias and the prediction of grades in law school (Law School Admission Research, LSAR-75-1. Law School Admissions Council, 1975, 1-52. Also in Journal of Legal Education, 1975, 27, 293-323. (b)

Linn, R. L., & Werts, C. E. Consideration for students of test bias. Journal of Educational Measurement, 1971, 8, 1-4.

Lunneborg, P. W., & Radford, D. The LSAT: A survey of actual practice. Journal of Legal Education, 1966, 18, 313-324.

Melden, A. I. Rights and persons. Berkeley & Los Angeles: University of California Press, 1977.

Mortimore, G. W. Rational action. In Benn & Mortimore (Eds.), Rationality and the social sciences. London: Routledge & Kegan Paul, 1976.

Nagel, T. Equal treatment and compensatory discrimination. Philosophy and Public Affairs, Summer 1973, 2(4), 348-363.

Novick, M. R., & Ellis, D. D., Jr. Equal opportunity in educational and employment selection. Paper presented at the annual meeting of the American Psychological Association, Chicago, 1975. Also in American Psychologist, 197[illegible] 32(5), 306-320.

Novick, M. R., & Petersen, N. S. Towards equlizing educational and employment opportunity. Journal of Educational Measurement, Spring 1976, 13(1), 77-88.

Ohio, Brief of the State of. Amicus Curiae in support of the respondents in the case of Defunis v. Odegaard, 461 U.S.

O'Neill, R. M. Discriminating against discrimination. Indianapolis: Indiana University Press, 1975.

Petersen, N. S., & Novick, M. R. An evaluation of some models for test bias. Iowa Testing Programs Occasional Papers. The American College Testing Program, September 1974, 8.

Petersen, N. S., & Novick, M. R. An evaluation of some models of culture-fair selection. Journal of Educational Measurement, Spring 1976, 13(1), 3-30.

Raiffa, H. Decision analysis: Introductory lectures on choice under uncertainty. Reading, Mass.: Addison-Wesley, 1968.

Rawls, J. Justice as fairness. In H. A. Bedau (Ed.), Justice and equality. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Rawls, J. A theory of justice. Cambrdige, Mass.: The Belknap Press of Harvard University Press, 1971.

Richards, D. A. J. A theory of reasons for action. Oxford: Clarendon Press, 1971.

Sawyer, R. L., Cole, N. S., & Cole, J. W. L. Utilities and the issue of fairness in a decision theoretic model for selection. Journal of Educational Measurement, Spring 1976, 13(1), 54-76.

Schmidt, F. L., & Hunter, J. E. Racial and ethnic bias in psychological tests: Divergent implications for two definitions of test bias. American Psychologist, 1974, 29, 1-8.

Sher, G. Justifying reverse discrimination in employment. Philosophy and Public Affairs, Winter 1975, 4(2), 159-170.

Simon, H. A. Models of man. New York: John Wiley & Sons, 1957.

Simon, R. Preferential hiring: A reply to Judith Jarvis Thomson. Philosophy and Public Affairs, Spring 1974, 3(3), 312-320.

Spencer, T. L. An investigation of the National Teacher Examination for bias with respect to black candidates. Doctoral dissertation, University of Colorado. Ann Arbor, Mich.: University Microfilms, 1972.

Stanley, J. C. Predicting college success of the educational disadvantaged. Science, 1971, 171, 640-647.

Strike, K. A. Justice and reverse discrimination. School Review, August 1976, 516-537.

Thomson, J. J. Preferential hiring. Philosophy and Public Affairs, Summer 1973, 2(4), 364-384.

Thorndike, R. L. Concepts of culture fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Tobriner, J. Dissenting opinion in the case of Bakke v. the Regents of the University of California, San Francisco, 23311.

Vlastos, G. Justice and equality In R. Brandt (Ed.), Social justice. Englewood Cliffs, N.J.: Prentice-Hall, 1962.

Von Neuman, J., & Morgenstern, O. Theory of games and economic behavior (2nd ed.). Princeton, N.J.: Princeton University Press, 1947.

Wald, A. Statistical decision functions. New York: John Wiley, 1950.

Wasserstrom, R. The university of the case for preferential treatment. American Philosophical Quarterly, April 1976, 13(2), _____).

Wasserstrom, R. Racism, sexism, and preferential treatment: An approach to the issues. UCLA Law Review, February 1977, 24(2), 581-622.

Williams, R. L. Problems of match and mismatch in testing black children. Paper presented at the annual meeting of the American Psychological Association, Honolulu, Hawaii, September 1972.

Yvette, S. F. The choice: The issue of black survival in America. New York: Berkeley Publishing Co., 1971.

1FSE:a

Footnotes

[1]I would like to thank Professor Robert L. Linn for his writings and discussions which have helped me in many ways.

[2]For all of these discussions I shall use the word "test" in its generalized sense to refer to the combination of predictors, which may include standardized tests and grade point averages.

[3]Two groups could have common prediction equations and still the accuracy of prediction is better for one group than the other. There would be no systematic errors if the regression equations were the same, but the standard error of estimate, however, would be larger for the one group than the other. I shall neglect such things here. See the section on the Equal Risk condition. It is interesting to note that ACT concluded that although there were differences in (multiple) correlations, the standard errors of estimate were comparable on the average for the majority and minority groups which were defined by racial-ethnic characteristics (Academic College Testing Assessment Program, p. 243). It has also been suggested that the lack of reliability in a test may make a majority group regression equation overpredict the performance of the minority group members rather than underpredict it. For further discussion of this and another difficulty, see R. L. Linn and C. E. Werts, 1971.

[4]There are complications here. Several key writers have argued that "rationality" is an essentially contested notion and that a defensible, semi-stipulative definition is needed. For further discussion see Ellett, "Decision Theory, Rationality, and Moral Justice," November, 1979.

[5]There are complications here, too. Several writers have argued that "moral justice" requires a defensible, semi-stipulative definition. See footnote 4.

[6]This issue and the policy recommendation are discussed in more detail in Ellett, "A More Fair Policy and the Limits of Fair Selection," April, 1979.

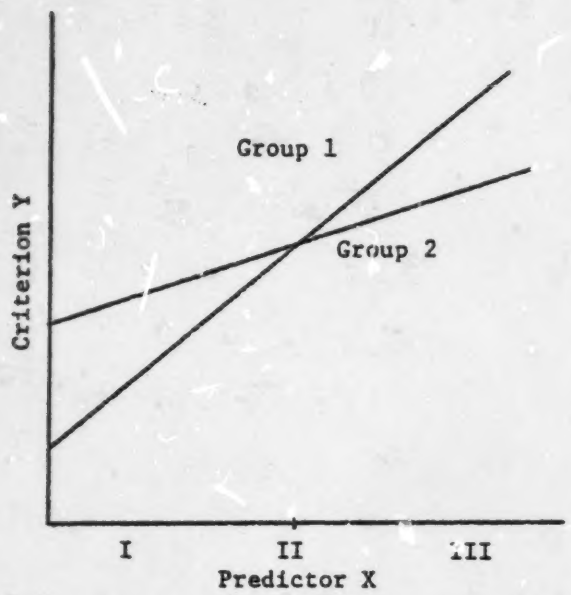

Figure 1a. Illustration of regression lines with unequal slopes

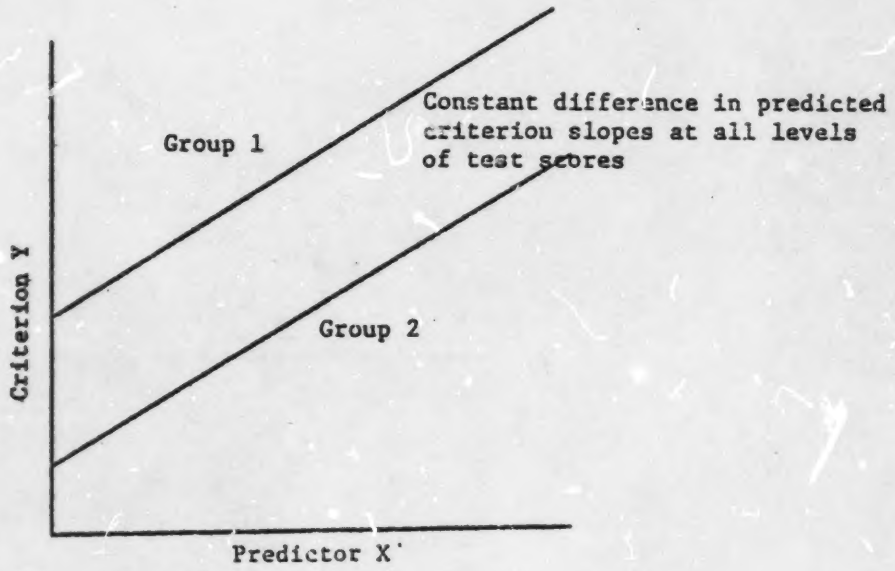

Figure 1b. Illustration of regression lines with equal slopes but unequal intercepts

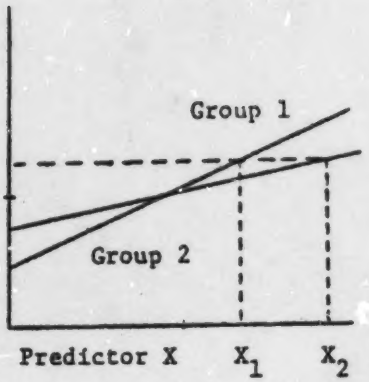

Figure 2a. The Regression Condition: Students with the highest predicted criterion scores, using separate equations within groups, are selected. In the graph above, a student in Group 1 with predictor score X_1 has the same predicted criterion score as a student in Group 2 with predictor score X_2. Thus, while the condition prescribes the selection of students predicted to do best in college, the example illustrates the case in which because prediction is poorer in one group (Group 2), members of that group with high predictor scores must score higher than members of another group to obtain the same predicted criterion score.

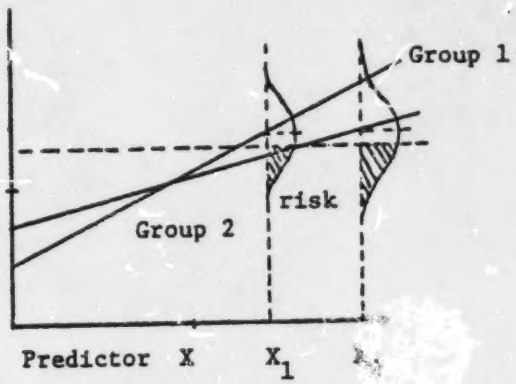

Figure 2b. The Equal Risk Condition: Students with the highest chances of success or smallest risk are selected. A student in Group 1 with predictor score X_1 has the same risk as a student in Group 2 with predictor score X_3. Thus, while the condition prescribes the selection of lowest risk students, the example illustrates the case in which because prediction is poorer in one group (Group 2), members of that group with highest predictor scores must score higher than members of another group to have the same risk.

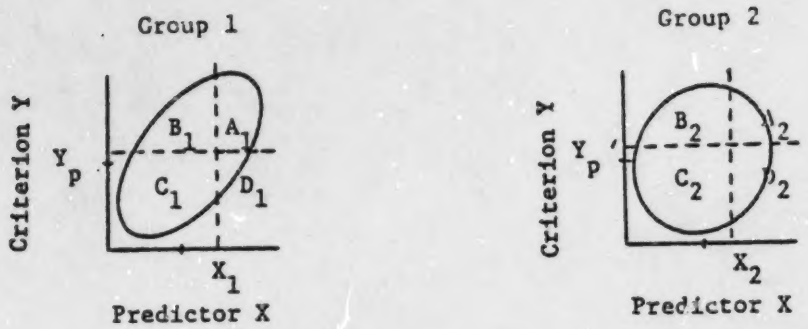

Figure 3a. The Constant Ratio Condition: Students are selected so that the ratio of the proportion selected to the proportion successful is the same in all groups. In the graph above an ellipse of the distribution of predictor and criterion scores is presented along with letters which represent the number of students falling in each of the four areas of the graph. Thus, the proportion selected is represented by (A+D)/(A+B+C+D) and the proportion successful by (A+B)/(A+B+C+D). X_1 and X_2 are set to satisfy the constant ratio model so that $(A_1+D_1)/(A_1+B_1)=(A_2+D_2)/(A_2+B_2)$. Although prediction is poorer in Group 2 than in Group 1, in contrast to the other conditions, X_2 is less than X_1. However, because a smaller proportion of Group 2 members are successful, members of that group have a smaller chance of selection, and very few of the potentially successful members of Group 2 are among those selected.

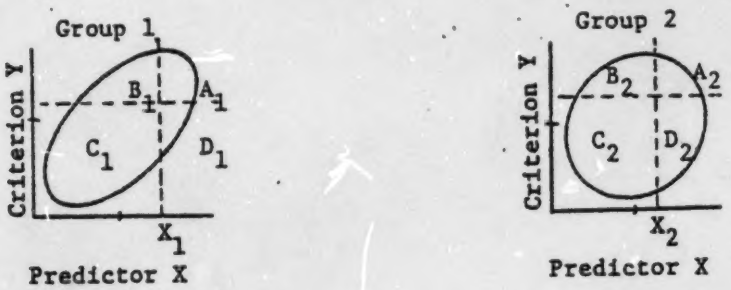

Figure 3b. The Conditional Probability Condition: Students are selected so that the conditional probability of selection given success is the same for all groups. In the graph above, A+B represents all successful students and therefore A/A+B represents the conditional probability of selection given success. X_1 and X_2 are set to satisfy the conditional probability model so that $A_1/(A_1+B_1) = A_2/(A_2+B_2)$. As with the constant ratio condition, although prediction is poorer in Group 2, X_2 is less than X_1.

		Criterion Performance Outcomes			
		Unsatisfactory	Satisfactory	Very Satisfactory	Exceptional
A.	Probability of Achieving Criterion	.20	.25	.30	.25
	Utility	0	.50	.60	.90
B.	Probability	.30	.35	.25	.10
	Utility	.10	.70	.80	1

Figure 4. Illustration of the expected utility strategy

ITSE:a